On the Modeling and Analysis of Sequential Observations of Spatial Processes with Application to Modern Earthwork Compaction

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Daniel Kristopher Heersink

aus den

Vereinigten Staaten von Amerika

Promotionskomitee

Prof. Dr. Reinhard Furrer (Vorsitz)
Prof. Dr. Leonhard Held

Zürich, 2013

To my mom.
You are no longer with us,
but you are always near to my heart.

ISAIAH 40:31

Preface

I owe my heartfelt thanks to many people for supporting me these past several years as I completed my PhD studies.

First I would like to thank Reinhard Furrer for providing me the opportunity to study with him and bringing me to Switzerland. I am grateful for the invaluable guidance and mentoring received. I am also thankful for the opportunity to travel to Enschede, Netherlands to attend a conference and to Golden, CO to work with Mike Mooney learning about rollers.

I would also like to thank Steve Geinitz for countless discussions as we worked on our PhDs together and for graciously accepting to review my thesis. I would also like to thank Mattia Molinaro for reviewing my thesis as well. A special thanks also goes to my various officemates that I have shared an office with these past three years. Thank you to Christina Bastian for proofreading my German.

Thank you to Mike Mooney and his team at Colorado School of Mines in Golden, CO for hosting me and explaining their research to me.

I would also like to thank Leonhard Held for being part of my dissertation committee and Paul Eilers for being my external evaluator.

Thank you to my father for being a great support my entire life. The biggest thanks goes to my wife Elizabeth for supporting me through this process and always being there for me and to Bronwyn for keeping every moment interesting. Thank you for your patience, understanding and love.

Financial support from the Swiss National Science Foundation is greatly acknowledged.

Zurich, February 2013

Daniel Heersink

Zusammenfassung

Räumlich-zeitliche Daten stellen viele herausfordernde Probleme dar, da sie die Umsetzung eines geeigneten stochastischen Modells und Schätzung Verfahren erfordern. Räumlich-zeitliche Daten werden häufig indirekt durch nichtlineare Measurement Operators und inhärent multivariaten mit komplexen Korrelation Strukturen beobachtet und in grossen Mengen erfasst. Ein Beispiel für solche Daten ist der Walze Messwert (RMV) welcher durch sequentielle moderne Erdverdichtungswalzen gesammelt wird.

Moderne Erdverdichtungswalzen sammeln einen nahezu kontinuierlichen Strom von Daten, die für die Qualitätssicherung (QA) und Qualitätskontrolle (QC) der Verdichtung verwendet werden können. RMVs können auch für intelligente Verdichtung (IC) verwendet werden, um durch die Anpassung der Betriebsparameter während der Bauphase eine homogene Verdichtung zu erreichen. Das Ziel des IC ist die Zeit der Verdichtung zu reduzieren und die Verdichtungsqualität durch die Identifizierung von Schwachstellen zu verbessern. Ein statistisches Modell des Bauplatzes ist erforderlich, um eine Software mit einer robusten Umsetzung zu bauen, um diese Schwachstellen "on the fly" zu identifizieren.

Zuerst wird ein stochastisches Modell von einer Verdichtungschicht entwickelt und zwei potentielle Schätzverfahren diskutiert. Das zuerst diskutierte Verfahren wird wahrscheinlich bestraft, da ein Glättungsparameter durch generalisierte Kreuzvalidierung gewählt wurde. Zweitens wird eine räumliche Backfitting Schätzverfahren vorgeschlagen. Backfitting ist ein iteratives-Schätzverfahren, wo fixe Effekte und zufällige Effekte wiederholt werden, bis Konvergenz der Schätzungen aktualisiert werden.

Die Untersuchung der komplexen Kovarianz Strukturen der vorgeschlagenen Modelle für RMVs führte zu der Entwicklung eine Verlängerung der Änderung vom Rang 1 der Pseudoinverses um die Rank Updates zu vergrössern. Die Pseudoinverse von $\mathbf{A} + \mathbf{X}_1\mathbf{X}_2^*$ wobei $\mathbf{A}, \mathbf{X}_1, \mathbf{X}_2$ sind komplexe Matrizen sind unter verschiedenen Annahmen gegeben. Wir nutzen das Ergebnis zur Ableitung der Pseudoinverse und inverse für die quasi-Kronecker strukturierten Matrix $\mathrm{bdiag}(\mathbf{A}_k) + \boldsymbol{u}\boldsymbol{v}^* \otimes \mathbf{E}$ mit p komplexen Matrizes $\mathbf{A}_k$ der Dimension $n \times m$, zwei komplexen p-Vektoren $\boldsymbol{u}$ und $\boldsymbol{v}$ und einer komplexen Matrix $\mathbf{E}$ der Dimension $n \times m$.

Wir schlagen als nächstes ein sequentielle, räumliche gemischtes Effektenmodell und ein sequentielles, räumliches Backfitting Verfahren zur Abschätzung der Modellierungsbegriffe vor. Die Schätzung der sequentiellen, räumlichen Prozesse ist recht komplex und mehrere Backfitting Verfahren werden vorgestellt unter Verwendung der quasi-Kronecker strukturierten Matrizen und die zuvor entwickelten Pseudoinverses.

Weiter werden die geschätzten Bereiche von der sequentiellen, räumliche Backfitting Verfahren hergestellt unter Verwendung einer Scale Space Multiskalenanalyse. Diese Bildanalyse wird als eine tragfähige Lösung für verbesserte IC und QA der Verdichtung für RMVs vorgeschlagen.

Schliesslich wird ein atypisch verdichteter Prüfstand von atypischen Abmessungen untersucht, um den Einfluss der Fahrtrichtung auf RMVs zu ermitteln. Explorative Analyse werden durchgeführt und empirische Semivariogramme geschätzt. Dann wird das sequentielle, räumliche Backfitting Verfahren auf die Daten angewendet, um die Bedeutung der Fahrtrichtung zu testen.

Abstract

Spatio-temporal data presents many challenging problems as they require implementation of a proper stochastic model and estimation procedure. Spatio-temporal data are often indirectly observed through non-linear measurement operators, inherently multivariate with complex correlation structures, and collected in huge quantities. An example of such data is the roller measurement value (RMV) sequentially collected by modern earthwork compaction rollers.

Modern earthwork compaction rollers collect a virtually continuous stream of data that can be used for quality assurance (QA) and quality control (QC) of the compaction process. RMVs can also be used for intelligent compaction (IC), adjusting operation parameters during construction to achieve homogeneous compaction. The goal of IC is to reduce compaction time and to improve compaction quality by identifying soft spots. A statistical model of the site is needed to build software with a robust implementation to identify these soft spots "on the fly".

First, a stochastic model of one compaction layer is developed and two potential estimation procedures are discussed. The first procedure discussed is penalized likelihood using a smoothing parameter chosen by generalized cross validation. Second, a spatial backfitting estimation procedure is proposed. Backfitting is an iterative estimation procedure where fixed effects and random effects are repeatedly updated until convergence of the estimates.

Investigating the complex covariance structures of proposed models for RMVs led to developing an extension of the rank one update of Moore–Penrose pseudoinverses to larger rank updates. The Moore–Penrose pseudoinverse of $\mathbf{A} + \mathbf{X}_1\mathbf{X}_2^*$, where $\mathbf{A}, \mathbf{X}_1, \mathbf{X}_2$ are complex matrices are given under various assumptions. We use the result to derive the Moore–Penrose pseudoinverse for the quasi-Kronecker structured matrix $\mathrm{bdiag}(\mathbf{A}_k) + \boldsymbol{u}\boldsymbol{v}^* \otimes \mathbf{E}$ with p complex matrices $\mathbf{A}_k$ of dimension $n \times m$, two complex p-vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ and a complex matrix $\mathbf{E}$ of dimension $n \times m$.

We next propose a sequential, spatial mixed-effects model and a sequential, spatial backfitting routine for estimation of the modeling terms. Estimation of sequential, spatial processes is quite complex and several backfitting routines are presented utilizing quasi-Kronecker structured matrices and the previously developed Moore–Penrose pseudoinverses.

Next, the estimated fields produced from the sequential, spatial backfitting procedure are analyzed using a multiresolution scale space analysis. This image analysis is proposed as a viable solution to improved IC and QA of the compaction process for RMVs.

Finally, an atypically compacted test bed of atypical dimensions is investigated to ascertain the influence of driving direction on RMVs. Exploratory analysis is performed and empirical semivariograms estimated. Then the sequential, spatial backfitting procedure is applied to the data to test the importance of driving direction.

Thesis outline

Introduction

Paper I: **Spatial Analysis of Modern Soil Compaction Roller Measurement Values**
Daniel K. Heersink & Reinhard Furrer
Conference Paper published in *Procedia Environmental Sciences* (2011), pp. 8–13.

Paper II: **On Moore–Penrose Inverses of quasi-Kronecker Structured Matrices**
Daniel K. Heersink & Reinhard Furrer
Paper published in *Linear Algebra and its Applications* (2012), 436, pp. 561–570.

Paper III: **Sequential Spatial Analysis of Large Datasets with Applications to Modern Earthwork Compaction Roller Measurement Values**
Daniel K. Heersink & Reinhard Furrer
Paper submitted to *Spatial Statistics*.

Paper IV: **Intelligent Compaction and Quality Assurance of Roller Measurement Values utilizing Backfitting and Multiresolution Scale Space Analysis**
Daniel K. Heersink, Reinhard Furrer, & Mike A. Mooney
Report, arXiv:1302.4631.

Paper V: **Spatial Backfitting of Roller Measurement Values from a Florida Test Bed**
Daniel K. Heersink, Reinhard Furrer, & Mike A. Mooney
Report, arXiv:1302.4659.

Introduction

1 Spatial Statistics

Often times, a researcher is concerned with the question of "how much is there?" Other times, of more importance can be the question of "how much is there and where is it located?" Spatial statistics addresses the second question by investigating spatial data (locations $\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n\}$ and data $\{Z(\boldsymbol{s}_1), \ldots, Z(\boldsymbol{s}_n)\}$ observed at those locations) and the variation within that data (Schabenberger and Gotway, 2005). The data is usually assumed to be random and the locations can be assumed fixed or random.

Spatial statistics is usually broken down into three categories, as detailed by Cressie (1993) – spatial processes indexed over a continuous space (geostatistical data), spatial processes indexed on a lattice (lattice data), and spatial point processes. This dissertation deals with geostatistical data.

1.1 Spatial Processes

A spatial stochastic process, or simply a spatial process, is a collection of random variables, with a well-defined joint distribution, indexed by a set $\mathcal{D} \subset \mathbb{R}^d$. The spatial process will be denoted as $\{Z(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{D} \subset \mathbb{R}^d\}$, where Z represents the attribute of interest that is observed and $\boldsymbol{s}$ is the location at which the observation is made. If attention is restricted to a finite set of spatial locations $\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n\} \subset \mathcal{D}$, then $(Z(\boldsymbol{s}_1), \ldots, Z(\boldsymbol{s}_n))^\mathsf{T}$ is a random vector with a multivariate distribution that reflects the spatial dependencies of the process (Gneiting and Guttorp, 2010). Most often, and in the case of this dissertation, $d = 2, 3$. For $d > 1$, a spatial process is termed a random field.

As an example, consider the concentration of mercury (Hg) in the soil. One intuitive characteristic of Hg concentrations is that a high concentration in a location $\boldsymbol{s}$ will probably mean a high concentration at points near $\boldsymbol{s}$. This is known as spatial autocorrelation as two points close together are more highly correlated than points further apart. The observations of this Hg concentration could be modeled by a random field.

One observation of a random field consists of a large (often huge) collection of observations at unique locations in $\mathcal{D}$. Multiple observations of the field requires multiple observations of the process at each spatial location. The difficulty in dealing with spatial processes is thus that there is often only one or a small number of observations of the field due to time and monetary constraints in collecting the data. Thus spatial variation needs to be ascertained from a small sample using the information contained in the spatial locations of the data (Schabenberger and Gotway, 2005).

The problem dealt with in spatial statistics is thus twofold. First, the data is assumed to have a complex correlation structure. This could potentially mean that every observation at a specific point is correlated with every observation at every other location. Secondly, the amount of data collected for each observation of the field can also present computational problems with memory and computing time. The covariance structure, i.e. the covariance matrix of the observations, increases as the square of the number of observations, e.g. has n^2 elements for n observations.

The spatial process can also be implemented into a spatio-temporal framework, see Cressie and Wikle (2011). Such a spatio-temporal framework includes dynamical and sequential processes. These processes not only have a spatial index, but they also have a temporal index.

Scales of Variation

Spatial data is often decomposed into an additive structure. The most basic decomposition can be symbolically represented as data = mean + error. The process $Z(\boldsymbol{s})$ can thus be modeled as a composition of a mean process and an error process: $Z(\boldsymbol{s}) = \mu(\boldsymbol{s}) + \varepsilon(\boldsymbol{s})$. The error process could be further decomposed into several sources of error. Let $\eta(\boldsymbol{s})$ be a small scale random process and $\xi(\boldsymbol{s})$ be a micro scale random process.

We could then model the process $Z(\boldsymbol{s})$ as

$$Z(\boldsymbol{s}) = \mu(\boldsymbol{s}) + \eta(\boldsymbol{s}) + \xi(\boldsymbol{s}).$$

In theory, this decomposition could encompass as many scales of variation as desired. The definitions of scale are unique to the modeler. One person's small scale variation, could be another person's micro scale variation (Cressie, 1993).

Gaussian vs. non-Gaussian Random Fields

Historically, random fields have not been assumed to be Gaussian. Thus, a large proportion of spatial statistics methods used in practice do not require a Gaussian assumption, i.e. methods developed in the fields of mining, hydrology, geography, etc. The assumption of Gaussian data leads to a much more straightforward analysis in many aspects. Unless specifically noted, Gaussianity will not be assumed in the following developments.

Stationarity

A random field is called a strict, or strong, stationary field if the spatial process is invariant under a translation of the coordinate system, i.e.

$$P(Z(\boldsymbol{s}_1) < z_1, \ldots, Z(\boldsymbol{s}_k) < z_k) = P(Z(\boldsymbol{s}_1 + \boldsymbol{h}) < z_1, \ldots, Z(\boldsymbol{s}_k + \boldsymbol{h}) < z_k),$$

for all k and $\boldsymbol{h}$ (Gneiting and Guttorp, 2010).

A relaxation of strict stationarity is termed second-order, or weak, stationarity. A random field is second-order stationary if the mean of the process is constant and the covariance between any two points in the field is only a function of the distance between those two points, i.e.

$E[Z(\boldsymbol{s})] = \mu$, for all $\boldsymbol{s}$, and $\mathrm{Cov}[Z(\boldsymbol{s}), Z(\boldsymbol{s}+\boldsymbol{h})] = C(\boldsymbol{h})$, for all $\boldsymbol{h}$. $C(\boldsymbol{h})$ is termed the covariance function. If the process is Gaussian, second-order stationarity implies strict stationarity (Schabenberger and Gotway, 2005).

If a spatial process is not second-order stationary, the increments $Z(\boldsymbol{s}) - Z(\boldsymbol{s}+\boldsymbol{h})$ might be. This is termed intrinsic stationarity. The process $\{Z(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{D} \subset \mathbb{R}^d\}$ is intrinsically stationary if $E[Z(\boldsymbol{s})] = \mu$ and

$$\frac{1}{2}\,\mathrm{Var}[Z(\boldsymbol{s}) - Z(\boldsymbol{s}+\boldsymbol{h})] = \gamma(\boldsymbol{h}), \tag{1.1}$$

where $\gamma(\boldsymbol{h})$ is termed the semivariogram of the spatial process (Schabenberger and Gotway, 2005).

Covariance Functions

A second-order stationary random spatial field is defined by a mean function and a covariance function,

$$C(\boldsymbol{h}) = \mathrm{Cov}[Z(\boldsymbol{s}), Z(\boldsymbol{s}+\boldsymbol{h})] = \mathrm{Cov}[Z(\mathbf{0}), Z(\boldsymbol{h})].$$

This covariance function has several properties (Schabenberger and Gotway, 2005):

- $C(\mathbf{0}) \geq 0$
- $C(\boldsymbol{h}) = C(-\boldsymbol{h})$
- $C(\mathbf{0}) \geq |C(\boldsymbol{h})|$
- If $C_j(\boldsymbol{h})$ are valid covariance functions, $j = 1, \ldots, k$, then $\sum_{j=1}^{k} b_j C_j(\boldsymbol{h})$ is a valid covariance function if $b_j \geq 0$ for all j.
- If $C_j(\boldsymbol{h})$ are valid covariance functions, $j = 1, \ldots, k$, then $\prod_{j=1}^{k} C_j(\boldsymbol{h})$ is a valid covariance function.
- If $C(\boldsymbol{h})$ is a valid covariance function in $\mathbb{R}^d$, then it is also a valid covariance function in $\mathbb{R}^p$, $p < d$.
- $C(\boldsymbol{h})$ is positive-definite, i.e. $\sum_{i=1}^{k}\sum_{j=1}^{k} a_i a_j C(\boldsymbol{s}_i - \boldsymbol{s}_j) \geq 0$
- $C(\boldsymbol{h})$ must be continuous for all $\boldsymbol{h} \neq \mathbf{0}$.

A random field $\{Z(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{D} \subset \mathbb{R}^d\}$ is said to be mean-square continuous at $\boldsymbol{s}$ if

$$\lim_{\boldsymbol{h}\to\mathbf{0}} \mathrm{E}[(Z(\boldsymbol{s}) - Z(\boldsymbol{s}+\boldsymbol{h}))^2] = 0.$$

This implies that the random field is mean-square continuous at $\boldsymbol{s}$ only if $C(\boldsymbol{h}) \to C(\mathbf{0})$ as $\boldsymbol{h} \to \mathbf{0}$ (Schabenberger and Gotway, 2005).

To aid in modeling, a parametric covariance function is usually assumed. $C(\boldsymbol{h})$ is potentially unique for every lag distance $\boldsymbol{h}$ between two locations in $\mathcal{D}$ and the assumption of a parametric function reduces the problem of identifying the covariance structure to identifying a small number of parameters.

A spatial process is termed isotropic if there is no second moment direction dependence, i.e. for a second-order stationary process, $C(\boldsymbol{h}) = C^*(\|\boldsymbol{h}\|)$, a function only of $\|\boldsymbol{h}\|$, where $\|\boldsymbol{h}\|$ is the Euclidean norm of the vector $\boldsymbol{h}$.

The Matérn class of isotropic covariance functions is very important for spatial statistics due to its wide use in applications. This class of covariance functions is parameterized as

$$C(\boldsymbol{h}) = \sigma^2 \frac{1}{\Gamma(\nu)} \left(\frac{\theta \|\boldsymbol{h}\|}{2} \right)^{\nu} 2K_\nu(\theta\|\boldsymbol{h}\|), \nu > 0, \theta > 0, \tag{1.2}$$

where K_ν is the modified Bessel function of the second kind of order ν, and $\Gamma(\nu)$ is the gamma function (Matérn, 1986).

For $\nu = 0.5$, we get what is termed the exponential covariance model:

$$C(\boldsymbol{h}) = \sigma^2 \exp(-\theta\|\boldsymbol{h}\|). \tag{1.3}$$

As $\nu \to \infty$, with appropriate scaling of θ, the Matérn covariance model is known as the Gaussian covariance model:

$$C(\boldsymbol{h}) = \sigma^2 \exp(-\theta\|\boldsymbol{h}\|^2). \tag{1.4}$$

The Gaussian covariance model is the limiting case of the Matérn covariance model and has very nice theoretical properties like infinite differentiability, but is not usually applicable to data since it is not likely to be found in natural phenomena (Cressie, 1993). Furthermore, Stein (1999) recommends not using the Gaussian covariance model as estimates of uncertainty using such a model are implausibly small for physical processes.

Another important isotropic covariance model is the spherical covariance model:

$$C(\boldsymbol{h}) = \begin{cases} \sigma^2(1 - 1.5(\|\boldsymbol{h}\|/\theta) + .5(\|\boldsymbol{h}\|/\theta)^3, & \|\boldsymbol{h}\| \leq \theta \\ 0, & \|\boldsymbol{h}\| > \theta \end{cases}, \tag{1.5}$$

where θ is the parameter governing the lag distance at which two points in $\mathcal{D}$ are no longer correlated (Chilès and Delfiner, 1999).

Semivariograms

The semivariogram of a random process is defined in equation (1.1). This is generally considered the crucial parameter of geostatistics, see Matheron (1963).

If the process is second-order stationary, then it is straightforward to show

$$\gamma(\boldsymbol{h}) = \frac{1}{2}\left\{2\sigma^2 - 2C(\boldsymbol{h})\right\} = C(\boldsymbol{0}) - C(\boldsymbol{h}), \tag{1.6}$$

where $\mathrm{Var}[Z(\boldsymbol{s})] = \mathrm{Var}[Z(\boldsymbol{s} + \boldsymbol{h})] = C(\boldsymbol{0}) = \sigma^2$ (Schabenberger and Gotway, 2005).

Important semivariogram models are the Matérn and spherical models, defined using equations (1.2), (1.5), and (1.6).

Under the assumption of second-order stationarity, the covariance function and the semivariogram thus contain the same information. Use of the semivariogram is more common though

as classical estimators of the semivariogram are unbiased, whereas the corresponding estimators of the covariance function are biased if the mean of the process must be estimated from the data (Cressie, 1993).

Nugget, Sill, and Range

The semivariogram describes the spatial dependence between two locations in $\mathcal{D}$ as a function of the distance (and possibly angle) between them. There are three distinct features of most semivariograms. If the semivariogram approaches a fixed value, either asymptotically or in a finite distance, and stays there for larger lag distances, this value reached is termed the total sill of the semivariogram. If the total sill is reached at a finite lag distance, the lag distance at which it is reached is termed the range of the semivariogram. If the total sill is only reached asymptotically, the value at which 95% of the total sill is reached is used and this is termed the practical range.

The semivariogram value $\gamma(\mathbf{0}) = 0$, but if $\gamma(\boldsymbol{h}) \to c_0 > 0$ as $\boldsymbol{h} \to \mathbf{0}$, c_0 is termed the nugget of the semivariogram (Cressie, 1993). Note that a positive value of the nugget implies the random field is not mean-square continuous at $\mathbf{0}$. The value of the total sill minus the nugget is termed the partial sill. See Figure 1.1 for a representative, theoretical spherical semivariogram.

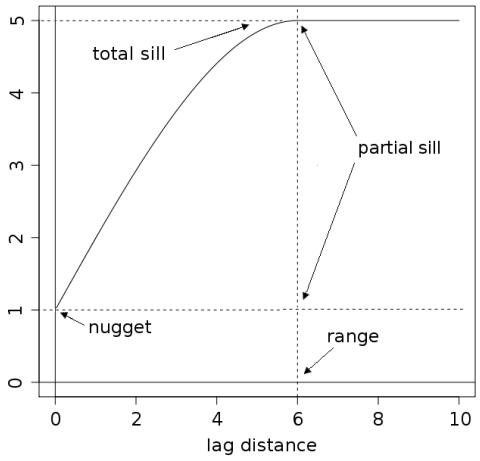

Figure 1.1: Spherical semivariogram with a nugget of 1, total sill of 5, and range of 6.

Oftentimes, the θ parameter in the exponential model (1.3) is re-parameterized as $\theta = 3/\alpha$, where α corresponds to the practical range of the model. For the Gaussian model (1.4), the re-parameterization used is $\theta = 3/\alpha^2$, where again, α corresponds to the practical range.

Semivariogram Estimation and Fitting

Analysis of data collected from a random field proceeds with the calculation and analysis of empirical semivariograms. These are constructed by binning the data into lag distances and estimating the value of γ for each bin.

The classical, empirical semivariogram estimator is the method-of-moments estimator, due to Matheron (1962), and termed the Matheron estimator:

$$\widehat{\gamma}(\boldsymbol{h}) = \frac{1}{2|N(\boldsymbol{h})|} \sum_{N(\boldsymbol{h})} \{Z(\boldsymbol{s}_i) - Z(\boldsymbol{s}_j)\}^2,$$

where $N(\boldsymbol{h})$ is the set of points that are a distance $\boldsymbol{h}$ (or possibly $\boldsymbol{h} \pm \varepsilon$) apart and $|N(\boldsymbol{h})|$ is the cardinality of this set. This estimator is sensitive to outliers and Cressie and Hawkins (1980) developed a robust estimator, termed the Cressie-Hawkins semivariogram estimator:

$$\tilde{\gamma}(\boldsymbol{h}) = \frac{1}{2} \left\{ \frac{1}{|N(\boldsymbol{h})|} \sum_{N(\boldsymbol{h})} |Z(\boldsymbol{s}_i) - Z(\boldsymbol{s}_j)|^{1/2} \right\}^4 \Bigg/ \left(0.457 + \frac{0.494}{|N(\boldsymbol{h})|} \right).$$

Empirical semivariograms can then be fitted to a theoretical semivariogram model. Estimation of the model parameters can be done using any of a number of estimation procedures including method of moments and maximum likelihood.

Anisotropy

Anisotropy occurs when the covariance function, or semivariogram, is direction dependent, i.e. points in a random field are more correlated in one direction than they are in another. Anisotropy can manifest itself in the nugget, sill, and/or the range of the semivariogram. Oftentimes a coordinate transformation will correct the anisotropy. This type of anisotropy is termed geometric anisotropy (Matérn, 1986).

Zimmerman (1993) argues that an experimental semivariogram that exhibits a direction dependent nugget effect may indicate that the measurement error is not adequately described by white noise. A sill anisotropy can be explained by a correlation of the measurement errors. Zimmerman (1993) also states that range anisotropy does not create any problems with the analysis of the semivariograms as the anisotropy can be handled with a transformation of the coordinate system.

1.2 Prediction (kriging)

Consider the classic spatial model

$$Z(\boldsymbol{s}) = \boldsymbol{x}(\boldsymbol{s})^{\mathsf{T}}\boldsymbol{\beta} + \boldsymbol{e}(\boldsymbol{s}), \qquad \boldsymbol{e}(\boldsymbol{s}) \sim (\mathbf{0}, \boldsymbol{\Sigma}(\boldsymbol{\theta})),$$

where $\boldsymbol{x}(\boldsymbol{s})^{\mathsf{T}}\boldsymbol{\beta}$ is a mean function and $\boldsymbol{e}(\boldsymbol{s})$ is a spatially correlated error process with $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ a positive definite covariance matrix parameterized by $\boldsymbol{\theta}$. Let $\boldsymbol{z}(\boldsymbol{s}) = (z(\boldsymbol{s}_1), \ldots, z(\boldsymbol{s}_n))^{\mathsf{T}}$ be observed data of the process $\{Z(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{D} \subset \mathbb{R}^d\}$, $\mathbf{X}$ be the matrix of fixed effects covariates, $\mathbf{X}$ of full rank. When we estimate $\boldsymbol{\theta}$ from the data, the estimate for $\boldsymbol{\beta}$ is the estimated generalized least squares (EGLS) estimator (Cressie, 1993):

$$\widehat{\boldsymbol{\beta}}_{egls} = (\mathbf{X}^{\mathsf{T}}\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathsf{T}}\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})^{-1}\boldsymbol{z}(\boldsymbol{s}).$$

When a prediction is desired for a new location $\boldsymbol{s}_0$, a method called kriging is done. Kriging was termed by Matheron (1963) after D. G. Krige, who developed empirical methods for determining ore-grade distributions from samples, see Krige (1951).

Kriging is a minimum mean squared error method of spatial prediction. Spatial prediction is the method of using information in data already gathered from a random process to predict the value of that process at an unobserved location.

If the mean of the process is known, spatial prediction is termed simple kriging. Consider the spatial data $\boldsymbol{z}(\boldsymbol{s}) = (z(\boldsymbol{s}_1), \ldots, z(\boldsymbol{s}_n))^\mathsf{T}$ and assume

$$\boldsymbol{z}(\boldsymbol{s}) = \boldsymbol{\mu}(\boldsymbol{s}) + \boldsymbol{e}(\boldsymbol{s}), \qquad \boldsymbol{e}(\boldsymbol{s}) \sim (\mathbf{0}, \boldsymbol{\Sigma}),$$

where $\boldsymbol{\mu}(\boldsymbol{s})$ and $\boldsymbol{\Sigma}$ are known. Here, $E[\boldsymbol{z}(\boldsymbol{s})] = \boldsymbol{\mu}(\boldsymbol{s})$ and $\mathrm{Var}[\boldsymbol{z}(\boldsymbol{s})] = \boldsymbol{\Sigma}$.

The best linear unbiased predictor (BLUP) at $\boldsymbol{s}_0$ is the simple kriging predictor:

$$p_{sk}(Z; \boldsymbol{s}_0) = \boldsymbol{\mu}(\boldsymbol{s}_0) + \boldsymbol{\sigma}^\mathsf{T}\boldsymbol{\Sigma}^{-1}(\boldsymbol{z}(\boldsymbol{s}) - \boldsymbol{\mu}(\boldsymbol{s})),$$

where $\boldsymbol{\sigma} = \mathrm{Cov}[\boldsymbol{z}(\boldsymbol{s}), z(\boldsymbol{s}_0)]$ (Schabenberger and Gotway, 2005). In the case of Gaussian data, this is *the* best predictor.

The associated simple kriging variance is

$$\sigma^2_{sk}(\boldsymbol{s}_0) = \sigma^2 - \boldsymbol{\sigma}^\mathsf{T}\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma},$$

where $\sigma^2 = \mathrm{Var}[z(\boldsymbol{s}_0)]$ (Cressie, 1993).

When the mean is unknown, but constant, i.e. $\boldsymbol{\mu}(\boldsymbol{s}) = \mu\mathbf{1}$, a procedure called ordinary kriging is used. The BLUP at $\boldsymbol{s}_0$ in this case is the ordinary kriging predictor:

$$p_{ok}(Z; \boldsymbol{s}_0) = \widehat{\mu} + \boldsymbol{\sigma}^\mathsf{T}\boldsymbol{\Sigma}^{-1}(\boldsymbol{z}(\boldsymbol{s}) - \mathbf{1}\widehat{\mu}),$$

where $\widehat{\mu}$ is the generalized least squares estimator:

$$\widehat{\mu} = (\mathbf{1}^\mathsf{T}\boldsymbol{\Sigma}^{-1}\mathbf{1})^{-1}\mathbf{1}^\mathsf{T}\boldsymbol{\Sigma}^{-1}\boldsymbol{z}(\boldsymbol{s}).$$

The associated ordinary kriging variance is (Cressie, 1993)

$$\sigma^2_{ok}(\boldsymbol{s}_0) = C(\mathbf{0}) - \boldsymbol{\sigma}^\mathsf{T}\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma} + \frac{(1 - \mathbf{1}^\mathsf{T}\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma})^2}{\mathbf{1}^\mathsf{T}\boldsymbol{\Sigma}^{-1}\mathbf{1}}.$$

Suppose that we have data $z(\boldsymbol{s}_1), \ldots, z(\boldsymbol{s}_n)$ and want to predict $z(\boldsymbol{s}_0)$ using the general linear model

$$\begin{aligned} \boldsymbol{z}(\boldsymbol{s}) &= \mathbf{X}(\boldsymbol{s})\boldsymbol{\beta} + \boldsymbol{e}(\boldsymbol{s}) \\ z(\boldsymbol{s}_0) &= \boldsymbol{x}(\boldsymbol{s}_0)^\mathsf{T}\boldsymbol{\beta} + \boldsymbol{e}(\boldsymbol{s}_0), \end{aligned}$$

where $\mathbf{X}(\boldsymbol{s})$ and $\boldsymbol{x}(\boldsymbol{s}_0)$ are, respectively, a matrix and vector of explanatory variables. Then the universal kriging predictor is defined as (Schabenberger and Gotway, 2005)

$$p_{uk}(Z; \boldsymbol{s}_0) = \boldsymbol{x}(\boldsymbol{s}_0)^\mathsf{T}\widehat{\boldsymbol{\beta}}_{gls} + \boldsymbol{\sigma}^\mathsf{T}\boldsymbol{\Sigma}^{-1}(\boldsymbol{z}(\boldsymbol{s}) - \mathbf{X}(\boldsymbol{s})\widehat{\boldsymbol{\beta}}_{gls}),$$

where $\widehat{\boldsymbol{\beta}}_{gls} = \left(\mathbf{X}(\boldsymbol{s})^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\mathbf{X}(\boldsymbol{s})\right)^{-1}\mathbf{X}(\boldsymbol{s})^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{z}(\boldsymbol{s})$, the generalized least squares estimate of $\boldsymbol{\beta}$. The kriging variance associated with this predictor is

$$\sigma^2_{uk}(\boldsymbol{s}_0) = \sigma^2 - \boldsymbol{\sigma}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma} + \left(\boldsymbol{x}(\boldsymbol{s}_0)^{\mathsf{T}} - \boldsymbol{\sigma}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\mathbf{X}(\boldsymbol{s})\right) \times \left(\mathbf{X}(\boldsymbol{s})^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\mathbf{X}(\boldsymbol{s})\right)^{-1}\left(\boldsymbol{x}(\boldsymbol{s}_0)^{\mathsf{T}} - \boldsymbol{\sigma}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\mathbf{X}(\boldsymbol{s})\right)^{\mathsf{T}}.$$

2 Roller Measurement Values (RMVs) and Engineering Goals

Modern earthwork compaction rollers are employed to compact material during road construction. These rollers collect compaction and location data as they operate. This data, termed the roller measurement value (RMV), can be modeled using a spatial process. An important goal of such a model is to improve the quality of the compaction process. This is achieved by improving the identification of weak, or soft spots, and by ensuring a homogeneous compaction.

2.1 History of RMVs

The first rollers designed for continuous compaction control (CCC) were used for construction starting in the 1970s in the European community. Rudimentary intelligent compaction (IC) technology was first available in the late 1990s. Dr. Heinz Thurner of the Swedish Highway Administration performed the first tests with an accelerometer mounted on a vibratory drum in 1974. This led to the introduction of the first RMV termed the compaction meter value (CMV) in 1978 (Mooney and Adam, 2007). The CMV is computed as the ratio of the amplitude of vertical drum acceleration at the first harmonic frequency to that of the base frequency of the vibrating drum:

$$CMV = C\frac{A_{2\Omega}}{A_{\Omega}},$$

where the C is a constant established during on site calibration and Ω is the operating frequency of the drum (Thurner and Sandström, 1980). Roller manufacturers employing the CMV include Dynapac and Caterpillar (Mooney and Adam, 2007).

Recently, the roller manufacturer Sakai has introduced a RMV called the continuous compaction value (CCV) that is based on the CMV. In addition to the base frequency and the first harmonic frequency, it utilizes information from first subharmonic frequency and also higher order harmonic frequencies (Mooney *et al.*, 2010). The CCV is computed as:

$$CCV = \left[\frac{A_{0.5\Omega} + A_{1.5\Omega} + A_{2.5\Omega} + A_{3\Omega}}{A_{2.5\Omega} + A_{3\Omega}}\right] \times 100.$$

In the 1990s, roller manufacturer Bomag introduced a more complex RMV termed the vibration modulus (E_{vib}), which provides a dynamic measure of soil stiffness by utilizing lumped parameter modeling and cylinder on elastic half-space theory (Kröber *et al.*, 2001). The drum/-soil assembly is modeled in Figure 2.1, where m_f and m_d are the masses of the frame and the drum respectively, g is the acceleration of gravity, z_d and $\ddot{z}_d$ are the vertical drum displacement

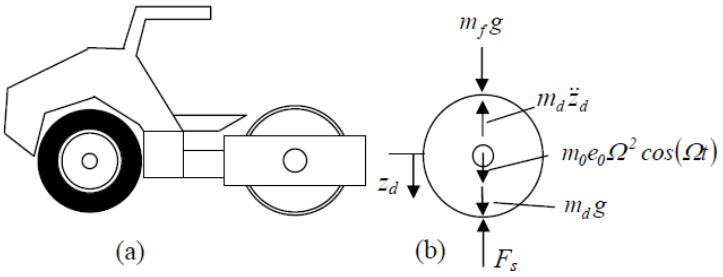

Figure 2.1: Model of drum/soil assembly for calculation of E_{vib}. Figure courtesy of Mooney and Adam (2007).

and acceleration, $m_0 e_0$ is the eccentric mass moment, Ω is the excitation frequency (32Hz in the case of Bomag), and F_s is the drum/soil contact force. Bomag employs two accelerometers to calculate z_d and F_s and from there calculates E_{vib} using Lundberg's theoretical solution for a rigid cylinder resting on a homogeneous, isotropic elastic half-space (Mooney and Adam, 2007).

The fourth RMV currently used in practice was developed by Ammann in the late 1990s. Ammann uses the same drum/soil assembly model as that depicted in Figure 2.1 to calculate the soil stiffness value (k_s) (Mooney *et al.*, 2010). Equating the forces of motion of the roller in the z-direction and solving for k_s, ignoring frame inertia, when the drum velocity is zero yields:

$$k_s(t) = \Omega^2 \left[m_d + \frac{m_0 e_0 \cos(\Omega t)}{z_d} \right]$$

(Anderegg and Kaufmann, 2004).

Quality assurance (QA) specifications for compaction using CCC were introduced in Europe in the 1990s. The most common use of CCC is to identify weak areas for evaluation via any of a number of spot testing procedures, including static plate load test (PLT) and light weight deflectometer (LWD), see Figure 2.2. Acceptance of these areas is based on the weak spots meeting prespecified criteria for the chosen spot testing (Mooney *et al.*, 2010).

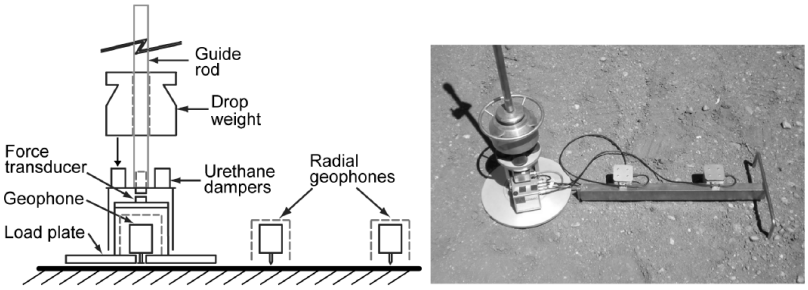

Figure 2.2: Schematic of LWD and picture of Prima 100 LWD. Figure courtesy of Senseney and Mooney (2010).

When the CMV was first introduced, vibration was achieved using a "clam shell" eccentric mass assembly inside the drum. Rotating the two eccentric masses in one direction with fre-

quency Ω provided maximum eccentric mass moment $m_0 e_0$ and maximal centrifugal force $F(t)$, defined as

$$F(t) = m_0 e_0 \Omega^2 \cos(\Omega t) = F_{ev} \cos(\Omega t),$$

where F_{ev} is the vertical excitation force (Mooney *et al.*, 2010). Operation in the opposite direction resulted in a minimum. In the 1990s, Bomag introduced the first roller capable of more than two amplitudes. Ammann and Dynapac have since introduced their own multiple amplitude vibrating rollers. The maximum vertical excitation force F_{ev} is commonly known as the theoretical amplitude A:

$$A = \frac{m_0 e_0}{m_d} = \frac{F_e v}{m_d \Omega^2},$$

and is the peak displacement of the drum (with mass m_d) if suspended in air (Mooney *et al.*, 2010).

The introduction of these newer rollers has led to some manufacturers implementing IC, where the eccentric excitation force is automatically controlled. If the roller completely loses contact with the soil, a state known as "jump mode", the force is decreased until contact is reattained. Bomag and Ammann/Case have also implemented IC where the excitation amplitude is decreased when a user defined threshold RMV is reached. This is a very rudimentary intelligence, but advances in theory and software are expected in the next years.

2.2 Uncertainty in RMVs

RMVs are calculated from frequency and excitation force amplitude information gathered from sensors within the vibrating and rotating smooth drum. All of the four RMVs discussed in Section 2.1 are directly proportional to the stiffness of the underlying soil (Facas and Mooney, 2011). The E_{vib} and k_s RMVs *directly* represent soil stiffness by calculating the contact force and displacement during vibration (Kröber *et al.*, 2001; Mooney and Rinehart, 2007). The CMV and CCV RMVs *indirectly* represent soil stiffness by giving a measure of the nonlinearity associated with the loss of contact during vibration (Thurner and Sandström, 1980; Scherocman *et al.*, 2007; Adam and Kopf, 2004). Many studies and models have verified that these RMVs represent soil stiffness, i.e. Yoo and Selig (1979), Adam and Kopf (2004), Anderegg and Kaufmann (2004), van Susante and Mooney (2008).

These RMVs reflect the aggregate stiffness of an approximately 1m^3 bulb of the underlying material, 2m long, 0.5m wide and approximately 1m in depth (Rinehart and Mooney, 2008). Due to this complication of multiple layers of material, the RMVs are often correlated to industry standard spot tests (White and Thompson, 2008). This correlation between RMVs and spot test measurements is somewhat suspect given the different measurement depths of the two devices (1m vs. 0.2–0.3m).

The bias of RMVs cannot be determined given the complex nature of the compaction process. Therefore, RMVs are a relative measure of soil stiffness (Mooney and Rinehart, 2009). This is not as hindering as it looks as many industry standard spot tests are likewise relative measures of soil stiffness, e.g., falling weight deflectometer, LWD, and soil stiffness gauge (Puppala, 2008). Therefore, the uncertainty needing quantification is that termed precision, the repeatability of measurements.

There are two values of uncertainty in RMVs. The uncertainty in a single RMV is σ_1 and the uncertainty in the difference between two different readings at the same spatial location is σ_2. Assuming that the error follows a Gaussian distribution with variance σ_1^2, the two uncertainties are related by

$$\sigma_2 = \sqrt{2}\sigma_1. \tag{2.1}$$

The empirical method used to estimate σ_2 is to compute the standard deviation of a difference vector of two passes through the compaction site. The data is gridded to ensure locations correspond (Facas and Mooney, 2011). This estimate is termed $\hat{\sigma_2}^D$ and equation (2.1) can be used to determine $\hat{\sigma_1}^D$

Variograms, and specifically the nugget, are used to estimate σ_1 in what is termed the nugget method. This estimate is denoted $\hat{\sigma_1}^N$ and $\hat{\sigma_2}^N$ can be determined by again using equation (2.1). Semivariogram analysis has been implemented on RMV data by Petersen *et al.* (2007), White and Thompson (2008), and Facas and Mooney (2011).

Due to the nature of roller movement through the compaction site, an empirical value of the nugget cannot be determined as the roller is always moving and there are no data points a zero distance apart. In practice, the first empirical semivariogram point is used as the value of the nugget for ease of determination. This value has been shown to be similar to the nugget estimation using a Gaussian covariance function and is thus a conservative estimate (Facas and Mooney, 2011).

2.3 Important Engineering and Modeling Questions

The standard practice of earthwork construction involves several layers of increasingly stiffer material being compacted on top of each other. The least stiff layer is the existing soil and is termed the (stabilized) subgrade. On top of this layer are multiple lifts, or layers, of sandy, low quality gravel of a greater stiffness than the subgrade. These layers are termed the subbase. The subbase is usually 0–1m thick in 15–30cm layers. The final layer of earthwork is a 15–30cm thick base layer composed of high quality gravel that is very stiff.

Since the sensors on the rollers measure to a depth of approximately 1m, the sensor reports an aggregate stiffness of multiple layers of earthwork of several possibly different stiffnesses. The physical modeling of this complex structure is in its infancy and only 1-dimensional. Intuitively, the sensor will report a weighted average of the stiffness of all layers that is much more heavily weighted on the less stiff layers.

Rollers can be used for QA. Currently QA is done by two methods. The first method uses spot test measurements. One test is performed over a volume of the area and if that spot passes, the whole area passes. The other method is proof rolling. Proof rolling involves a heavy wheeled vehicle that is driven over the test area. Passing or failing of the area is determined by observation of deflection under the wheels (Facas, 2009).

Rollers can also be used for IC. Some rollers are equipped with feedback control systems that allow changes to their operating parameters in real time. In general, when high RMVs are recorded the vibration amplitude is lowered and vice versa for low RMVs. The goal of feedback control of roller operating parameters is to reduce compaction time or to improve compaction

quality (Facas, 2009). A statistical model of the site is needed to build software to identify these soft spots.

The sensors employed for measuring soil properties are mounted near an end of the vibrating drum. This necessitates the inclusion of a driving direction covariate into any model of RMVs. The operator of the roller does not necessarily drive in a straight line. This leads to issues of independence as RMVs in consecutive lanes are not necessarily independent of one another. The issue is compounded when multiple layers are considered. Each lane of the compaction area usually does not lie directly on top of the lane below it. There tends to be a shifting in the transverse direction of the lanes from layer to layer. This shifting is not predictable and is potentially different for each successive layer. These are important things to consider when building a stochastic model.

RMVs are an aggregate measurement of the underlying soil stiffness. This aggregation is a weighted average of the bulb of soil directly beneath the drum that is dependent on the placement of the sensor within the drum. The vibrating, rotating drum also rotates about a y-directional centroid that is dependent on the y-directional heterogeneity of the underlying soil structure. Stochastic models of RMVs should take this behavior into account.

A model of each layer of the construction process is developed in Heersink and Furrer (2011). Investigations of the importance of driving direction are found in Heersink *et al.* (2013b). A model for multiple layers is developed in Heersink and Furrer (2013) and a proposal for QA/QC and IC is developed in Heersink *et al.* (2013a).

3 Aggregation Investigations

RMVs are data collected at points in space, but are not values of a point supported spatial process. Rather, the values are a complicated nonlinear aggregation of the random process. This aggregation is over a bulb of material approximately 2m long, 0.5m wide, and 1m deep. This aggregation also introduces correlation as each measurement bulb has some amount of overlap with the next measurement bulb (Rinehart and Mooney, 2008).

When inference is needed on aggregated, or averaged, processes whose supports are different than the data, this is known in the literature as a change of support problem (COSP) (Cressie, 1993).

Of interest in the study of this aggregation is whether properties of the spatial process can be inferred from the aggregated measurements of that process. To investigate the effect of this aggregating process, the aggregation was modeled as an unweighted integration over a volume. For a point supported spatial process $Z(\boldsymbol{s})$, the aggregate process $Z(B)$ is the average over all points in the volume $B : Z(B) = \frac{1}{|B|}\int_B Z(\boldsymbol{s})d\boldsymbol{s}$, where $|B|$ is the volume of B (Cressie, 1993).

Behavior of the semivariogram of the aggregate process is of interest. Behavior near the origin informs the differentiability of the process. Understanding this aggregation is the first step toward a more complicated model where a nonlinear weight function could be used. Establishment of a link between the semivariogram of the aggregated process to that of the true process allows inference of the true process.

3.1 Numerical Integration

The investigation of this COSP was done using the Matérn class of covariance functions. Three values of ν were chosen, 0.5, 1.5, and 2.5, to provide differing behavior at the origin. A value of 0.5 is the exponential covariance function and is linear at the origin. Values of 1.5 and 2.5 are both quadratic at the origin, but a value of 2.5 is more differentiable in the mean-squared error sense (Stein, 1999).

Simplistically said, in a polynomial expansion of the covariance function, the non-even powered terms govern the behavior and are therefore of greatest interest in the analysis (Stein, 1999). The parameters θ and σ^2 govern the range and the partial sill respectively. Both were chosen to be 1 for the study without loss of generality. Figure 3.1 is a plot of the three functions.

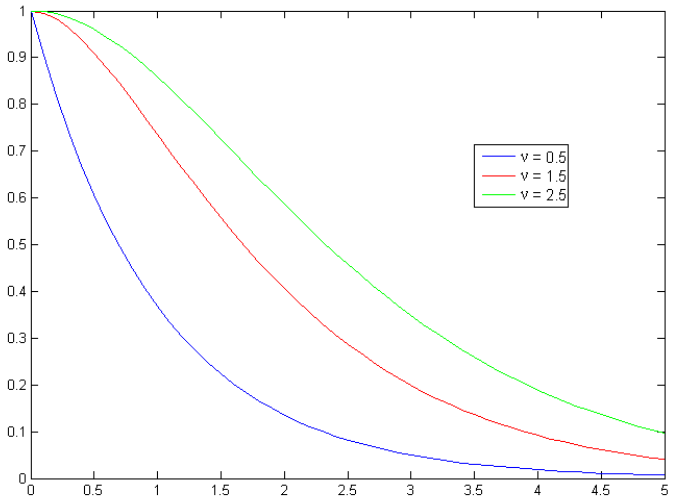

Figure 3.1: Matérn covariance functions, ν = 0.5, 1.5, 2.5

As the aggregation process is quite complex and is an average measurement over a volume that is not clearly defined, the first step in the integration is to define B. For this investigation, three choices of dimension of B were used: $B \in \mathbb{R}$, $B \in \mathbb{R}^2$, and $B \in \mathbb{R}^3$. Practically, only $\mathbb{R}^3$ is of interest as the roller measures a 3-dimensional volume. All three are investigated though to understand how the aggregation effect changes with dimension size. Lower dimensional problems are also more tractable for analytical analyses. For $\mathbb{R}$, B was chosen to be the line segment $[0, 1]$. For $\mathbb{R}^2$, B was chosen to be $[0, 1] \times [0, 2]$, and for $\mathbb{R}^3$, B was chosen as $[0, 1] \times [0, 2] \times [0, 1]$.

As the semivariogram in the driving direction is of interest, integration was only done over that direction. One point was held fixed and the other point was moved through the driving (x)-direction, simulating the movement of the roller. A plot of the Matérn function with $\nu = 1.5$ integrated over 1, 2, and 3 dimensions is provided in Figure 3.2. The integration over $\mathbb{R}^3$, $\int_B C(\|\boldsymbol{h}\|)\, d\boldsymbol{h}$, becomes

$$\int_0^2 \int_0^1 \int_0^1 \int_0^2 \int_0^1 \int_h^{1+h} C(\|\boldsymbol{u} - \boldsymbol{v}\|)\, du_1\, du_2\, du_3\, dv_1\, dv_2\, dv_3,$$

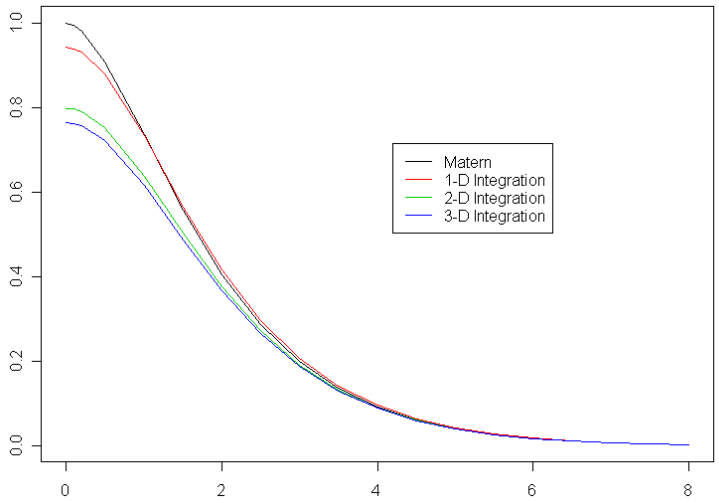

Figure 3.2: Matérn $\nu = 1.5$, Integrated on $\mathbb{R}$, $\mathbb{R}^2$, and $\mathbb{R}^3$

where $\boldsymbol{u}$ and $\boldsymbol{v}$ are any two points in B. A closed form of this integral is unattainable.

An analytic approximation to the integral is also not what is needed for the investigation. In practice, the shape of the covariance function must be estimated from the data. This estimation is done at distinct points, thus we have a noisy estimate of a curve at several points and never observe the true curve. A numerical integration provides the necessary information for this investigation.

All of the aggregated covariance functions exhibit three characteristics. They start at values less than the original semivariogram, cross the original function, and then asymptotically approach zero.

The first two of these characteristics can be explained by the averaging nature of the aggregation. As the aggregated process averages all of the point values within the volume B, there is less variability in the covariance function at small distances. As the distance between points is increased, this averaging process serves to slow the rate at which the variance decreases. As the rate of decrease of the variance slows, the two plots will cross and the aggregated function will have more variance than the original one.

The third characteristic is found in all covariance functions: as $\|\boldsymbol{h}\| \to \infty$, all covariance functions approach zero.

3.2 Origin Investigation

The next step in the analysis of the aggregation process is to investigate the behavior of the aggregated covariance at the origin. The data is assumed to come from an aggregated process, so the empirical semivariograms should have the same behavior at the origin as that of the aggregated semivariograms from the numerical integration.

One would like the origin behavior of the integrated covariance to be similar to the origin

behavior of the empirical semivariogram. This then would give credence to the integration model of the point supported spatial process. The pairs $(h, C_B(h))$ of the numerical integration were used to perform a fifth degree polynomial regression.

The (scaled) polynomial fits for the Matérn function integrated over $\mathbb{R}^3$ are:

$$\begin{aligned} 1 - 0.002x - 0.485x^2 + 0.002x^3 + 0.189x^4 - 0.014x^5 \text{ for } \nu = 0.5, \\ 1 - 0.233x^2 - 0.001x^3 + 0.054x^4 - 0.014x^5 \text{ for } \nu = 1.5, \text{ and} \\ 1 - 0.135x^2 - 0.001x^3 + 0.018x^4 - 0.006x^5 \text{ for } \nu = 2.5. \end{aligned}$$

For comparison, the Taylor expansion of the Matérn covariance function with $\nu = 0.5, 1.5, 2.5$ are

$$\begin{aligned} C_{0.5} &\approx 1 - x + \frac{1}{2}x^2 - \frac{1}{6}x^3 + \frac{1}{24}x^4 - \frac{1}{120}x^5, \\ C_{1.5} &\approx 1 - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \frac{1}{8}x^4 + \frac{1}{30}x^5, \text{ and} \\ C_{2.5} &\approx 1 - \frac{1}{6}x^2 + \frac{1}{24}x^4 - \frac{1}{45}x^5, \text{ respectively.} \end{aligned}$$

Comparing the polynomial fit to the Taylor expansion for $\nu = 0.5$, the aggregated process is approximately a Matérn covariance with $1.5 < \nu < 2.5$. Similar comparisons can be found for $\nu = 1.5, 2.5$.

Estimation of the ν parameter in the Matérn family of covariance functions is difficult. In this aggregation investigation, using simple, known volumes B, identifiability of the ν parameter is impossible. This identifiability is even more difficult if a more physically accurate volume were chosen. Very little information about the properties of the true process can be ascertained from the aggregated process. Therefore, the spatial process under investigation, while properly defined as $Z(B_i)$, will be referred to as $Z(\boldsymbol{s}_i)$ for convenience.

4 Three Potential Approaches

Let the RMVs be modeled as $\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{U}\boldsymbol{\alpha}_1 + \mathbf{V}\boldsymbol{\alpha}_2 + \boldsymbol{\varepsilon}$, where $\mathbf{X}$ is a full rank matrix of explanatory variables, $\boldsymbol{\beta}$ the unknown coefficients associated with the explanatory variables, $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$ are uncorrelated Gaussian spatial processes of large and micro scale, $\mathbf{U}$ and $\mathbf{V}$ are full rank, and $\boldsymbol{\varepsilon}$ is an uncorrelated, zero mean noise process, uncorrelated with $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2$.

Gaussian spatial processes are assumed as the RMVs are continuous and empirically follow a Gaussian distribution. Three potential estimation procedures for such a model are penalized likelihood utilizing a generalized cross validation approach, a full Bayesian approach with prior distributions on all of the variance parameters, and a spatial backfitting approach that converges to the estimated generalized least squares estimate by iteratively estimating the fixed effects and the spatial terms.

4.1 Splines and Generalized Cross Validation

Splines are a very general, and quite flexible data smoothing method. Data smoothing is a nonparametric approach to estimating a trend in data. A great deal of research into splines has been done and they have been applied to a wide range of data, i.e. Wahba (1990), Eilers and Marx (2004), Marx *et al.* (2011), Eilers (2003).

A thin-plate-spline (TPS) is a nonparametric smoothing approach. Consider the problem of estimating a smooth function $f(\boldsymbol{s})$ from n observations $(y_i, \boldsymbol{s}_i)$. Estimation of the TPS is the problem of finding the function $\widehat{f}$ that minimizes

$$\|\boldsymbol{y} - \mathbf{f}\|^2 + \lambda J_{md}(f),$$

where $\mathbf{f} = (f(\boldsymbol{s}_1), \ldots, f(\boldsymbol{s}_n))^\mathsf{T}$ is a smooth function, J_{md} is a penalty functional measuring the "wiggliness" of f and λ is the smoothing parameter controlling the balance between data fitting and smoothness. J_{md} is defined as (Wood, 2006)

$$J_{md} = \int \ldots \int_{\mathbb{R}^d} \sum_{\nu_1 + \cdots + \nu_d = m} \frac{m!}{\nu_1! \ldots \nu_d!} \left(\frac{\partial^m f}{\partial x_1^{\nu_1} \ldots \partial x_d^{\nu_d}} \right)^2 dx_1 \ldots dx_d.$$

This problem is equivalent to minimizing

$$\|\boldsymbol{y} - \mathbf{A}\boldsymbol{\alpha} \ - \mathbf{E}\boldsymbol{\delta}\|^2 + \lambda \boldsymbol{\alpha}^\mathsf{T} \mathbf{A} \boldsymbol{\alpha}, \tag{4.1}$$

with respect to $\boldsymbol{\alpha}, \boldsymbol{\delta}$, subject to $\mathbf{E}^\mathsf{T}\boldsymbol{\alpha} = \mathbf{0}$ (Wood, 2006). Here, $\mathbf{E}$ is the matrix containing the spline basis functions in the null space of J_{md}. For example, if $m = d = 2$ the basis functions are $\phi_1(\boldsymbol{s}) = 1, \phi_2(\boldsymbol{s}) = s_1, \phi_3(\boldsymbol{s}) = s_2$. $\mathbf{A}$ is the matrix containing the remaining basis functions

$$\phi_{md}(r) = \begin{cases} \frac{(-1)^{m+1+d/2}}{2^{2m-1}\pi^{d/2}(m-1)!(m-d/2)!} r^{2m-d} \log(r) & d \text{ even} \\ \\ \frac{\Gamma(d/2-m)}{2^{2m}\pi^{d/2}(m-1)!} r^{2m-d} & d \text{ odd.} \end{cases}$$

Model Setup

Let $\boldsymbol{\alpha}_1$ represent a large scale smoothing (thin plate) spline, and $\boldsymbol{\alpha}_2$ represent a micro scale (zero mean Gaussian) process. Smoothing of the field can then be done using a penalized likelihood approach where the optimal smoothing parameters are chosen using generalized cross validation. Penalized regression is a common method employed in a wide range of applications, i.e. Eilers (1991) and Perperoglou and Eilers (2010), and in the context of regression splines: Eilers *et al.* (1996) and Marx and Eilers (1998).

Let $\lambda_1, \lambda_2 > 0$ be the smoothing parameters for $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$ respectively. The goal is to find

$$\min_{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\beta}} \|\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{U}\boldsymbol{\alpha}_1 - \mathbf{V}\boldsymbol{\alpha}_2\|^2 + \lambda_1 {\boldsymbol{\alpha}_1}^\mathsf{T} \mathbf{H}_1 \boldsymbol{\alpha}_1 + \lambda_2 {\boldsymbol{\alpha}_2}^\mathsf{T} \mathbf{H}_2 \boldsymbol{\alpha}_2,$$

where $\mathbf{H}_1$ is symmetric, semi-positive definite, and $\mathbf{H}_2$ is the covariance matrix of $\boldsymbol{\alpha}_2$.

By letting $\mathbf{W} = [\mathbf{X}\ \mathbf{U}\ \mathbf{V}]$, $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2)^\mathsf{T}$, and $\boldsymbol{\Omega} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \mathbf{H}_1 & 0 \\ 0 & 0 & \frac{\lambda_2}{\lambda_1}\mathbf{H}_2 \end{pmatrix}$, we can rewrite this minimization as $\min_{\boldsymbol{\theta}} \|\boldsymbol{y} - \mathbf{W}\boldsymbol{\theta}\|^2 + \lambda_1 \boldsymbol{\theta}^\mathsf{T}\boldsymbol{\Omega}\boldsymbol{\theta}$. This is a well known minimization problem.

The hat matrix, $\mathbf{A}(\lambda_1, \lambda_2)$, is $\mathbf{W}(\mathbf{W}^\mathsf{T}\mathbf{W} + \lambda_1\boldsymbol{\Omega})^{-1}\mathbf{W}^\mathsf{T}$. The trace of $\mathbf{A}(\lambda_1, \lambda_2)$ also gives us an estimate of the degrees of freedom, or effective number of parameters, in the model.

Let $\mathbf{C}$ be a matrix such that $\mathbf{C}\mathbf{W}^\mathsf{T}\mathbf{W}\mathbf{C} = \mathbf{I}$ and let $\mathbf{B}$ be a matrix such that $\mathbf{C}\boldsymbol{\Omega}\mathbf{C} = \mathbf{B}\mathbf{D}\mathbf{B}^\mathsf{T}$, where $\mathbf{D}$ is a diagonal matrix and $\mathbf{B}\mathbf{B}^\mathsf{T} = \mathbf{I}$. Thus $\mathbf{D}$ contains the eigenvalues of $\mathbf{C}\boldsymbol{\Omega}\mathbf{C}$.

This implies that $\mathbf{A}(\lambda_1, \lambda_2) = \mathbf{W}\mathbf{C}\mathbf{B}(\mathbf{I} + \lambda_1\mathbf{D})^{-1}\mathbf{B}^\mathsf{T}\mathbf{C}^\mathsf{T}\mathbf{W}^\mathsf{T}$. This leads to

$$\mathrm{tr}(\mathbf{A}(\lambda_1, \lambda_2)) = \mathrm{tr}((\mathbf{I} + \lambda_1\mathbf{D})^{-1}) = p + \sum_{i=1}^{q} \frac{1}{1 + \lambda_1\Lambda_{H_1,ii}} + \sum_{j=1}^{r} \frac{1}{1 + \lambda_2\Lambda_{H_2,jj}},$$

where $\mathbf{D} = \mathrm{bdiag}(\mathbf{0}, \boldsymbol{\Lambda}_{H_1}, \boldsymbol{\Lambda}_{H_2})$, p is the dimension of $\mathbf{I}$, and q and r are the ranks of $\mathbf{H}_1$ and $\mathbf{H}_2$ respectively. Thus the smoothing parameters are identifiable.

The generalized cross validation equation is

$$GCV(\lambda) = \frac{1}{n}\sum_{i=1}^{n} w_i^2 \left(\frac{(y_i - \hat{y}_i)}{1 - \mathbf{A}(\lambda)_{ii}}\right)^2.$$

Therefore, if we let $w_i = \frac{1-\mathbf{A}(\lambda)_{ii}}{1-\mathrm{tr}(\mathbf{A}(\lambda))/n}$, then

$$GCV(\lambda) = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n\left(1 - \frac{1}{n}\left(p + \sum_{i=1}^{q}\frac{1}{1+\lambda_1\Lambda_{H_1,ii}} + \sum_{j=1}^{r}\frac{1}{1+\lambda_2\Lambda_{H_2,jj}}\right)\right)^2}. \tag{4.2}$$

Simulation Study

Data was simulated on a rectangular grid $\boldsymbol{x}$ from the model $\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{U}\boldsymbol{\alpha}_1 + \mathbf{V}\boldsymbol{\alpha}_2 + \boldsymbol{\varepsilon}$, where $\mathbf{X} = (\mathbf{1}, \boldsymbol{x})$, $\boldsymbol{\beta} = (1, 1, 1)^\mathsf{T}$, $\mathbf{U} = \mathbf{V} = \mathbf{I}_n$, $\boldsymbol{\alpha}_2$ is a zero mean exponential random field, $\mathbf{H}_1$ is the matrix of radial basis functions, i.e. $\|\boldsymbol{x}\|^2\log(\|\boldsymbol{x}\|)$, and $\boldsymbol{\varepsilon}$ is white noise. The data was then smoothed using the penalized likelihood approach from above.

The GCV was then calculated using equation (4.2) on a grid of λ_1 and λ_2 values. The minimum on the grid, $\boldsymbol{\lambda}_{min} = (\lambda_{1,min}, \lambda_{2,min})^\mathsf{T}$, was recorded for many realizations of the simulated data. The simulation was repeated for several different values of the signal to noise ratio for both $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$.

The minimum smoothing parameters tend to cluster at their true (theoretical) values. There is, however, a distinct clustering of values for some of the simulated fields at the boundary of the $\boldsymbol{\lambda}$ grid, see Figure 4.1. The limits of the grid were expanded to include larger $\boldsymbol{\lambda}$ values, clustering points remained on the boundary. This may be due to numerical precision problems as the GCV function never reaches a minimum.

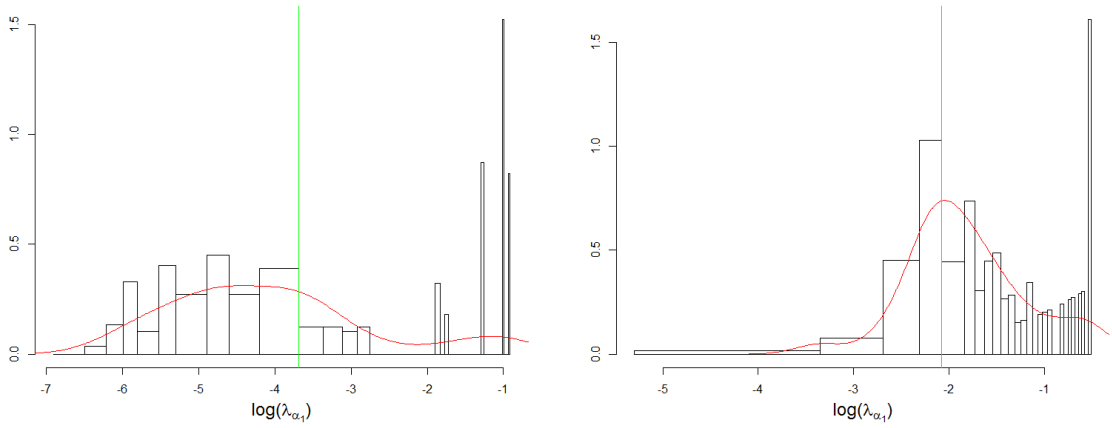

Figure 4.1: Histogram of $\log(\lambda_{\alpha_1})$ (left) and $\log(\lambda_{\alpha_2})$ (right). True value of the parameters is represented by a vertical green line.

To investigate this boundary issue, data was simulated on the same grid after removing one of the three modeling elements from the model. The elimination of the $\mathbf{X}\boldsymbol{\beta}$ term from the model does not eliminate the boundary clustering. This leads to the conclusion that the mean term in the model does not affect the GCV calculation or the boundary issues associated with the minimization.

Elimination of the $\mathbf{U}\boldsymbol{\alpha}_1$ element from the model successfully eliminates the boundary issues. This is an expected result as the model has been simplified to a constant mean term and an exponential random field with noise.

Elimination of the $\mathbf{V}\boldsymbol{\alpha}_2$ element from the model results in the same successful elimination of boundary issues. This simplified model is a constant mean term and a Matérn random field with noise. Both of these simplifications lead to well known problems and the results, as expected are consistent with theory.

The grid of $\boldsymbol{\lambda}$ values was then collapsed to one dimension and the λ_2 value held constant at the true value. The resultant marginal GCV function for λ_1 still displayed evidence of the boundary issues when minimizing in two dimensions. Calculating the marginal GCV while holding λ_1 constant at its true value still produces boundary issues when minimizing the λ_2 values.

Results

There are some simulated datasets that do not have a minimum GCV. To further investigate this problem, the spline part of the model was replaced with a Gaussian field with a Matérn covariance function and a range comparable to that of the spline structure. The smoothing parameter of this field, ν, was set to be 1.5. Results of this investigation indicate the boundary issues are not related to the spline construction as there are still fields generated with no minimum GCV. The boundary issues are potentially a numerical issue.

Beyond the boundary problems, most values of $\boldsymbol{\lambda}_{min}$ seem to follow a bivariate Normal distri-

bution centered at or very near the true value. The scale of the variance in either the $\boldsymbol{\alpha}_1$ or the $\boldsymbol{\alpha}_2$ direction is relative to the size of the true value of the parameter. For small values of the true parameter, the variance is smaller and vice versa.

4.2 Bayes Formulation

Using a hierarchical Bayesian model framework, we model $\boldsymbol{y}$ conditionally as a multivariate Gaussian:

$$\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \sigma_\varepsilon^2 \sim \mathcal{N}(\mathbf{X}\boldsymbol{\beta} + \mathbf{U}\boldsymbol{\alpha}_1 + \mathbf{V}\boldsymbol{\alpha}_2, \sigma_\varepsilon^2\mathbf{I}).$$

We then model

$$\boldsymbol{\alpha}_1|\mathbf{K}, \sigma_1^2 \sim \mathcal{N}(\mathbf{0}, \sigma_1^2\mathbf{K}^+) \text{ and} \tag{4.3}$$
$$\boldsymbol{\alpha}_2|\mathbf{G}, \sigma_2^2 \sim \mathcal{N}(\mathbf{0}, \sigma_2^2\mathbf{G}^+), \tag{4.4}$$

where $\mathbf{K}$ and $\mathbf{G}$ are symmetric semi-positive-definite hyperparameters. Note that $\mathbf{A}^+$ is the Moore–Penrose pseudoinverse of the matrix $\mathbf{A}$. If the matrix is positive definite, then $\mathbf{A}^+ = \mathbf{A}^{-1}$. Note that the probability density functions of these distributions only exists in the case of $\mathbf{K}$ and $\mathbf{G}$ positive-definite. This representation is thus for convenience of notation.

We then give the variance parameters inverse-gamma prior distributions:

$$\begin{aligned}
\sigma_1^2 &\sim \text{I}\Gamma(\eta, \nu) \\
\sigma_2^2 &\sim \text{I}\Gamma(\lambda, \psi), \text{ and} \\
\sigma_\varepsilon^2 &\sim \frac{1}{\sigma_\varepsilon},
\end{aligned}$$

where σ_ε^2 is given an improper prior. With these prior distributions, the posterior distributions of the variance parameters are straightforward to find analytically:

$$\begin{aligned}
\sigma_1^2|\boldsymbol{\alpha}_1, \ldots &\sim \text{I}\Gamma\left(\eta + \frac{\text{rank}(\mathbf{K})}{2}, \nu + \frac{1}{2}\boldsymbol{\alpha}_1{}^\mathsf{T}\mathbf{K}\boldsymbol{\alpha}_1\right) \\
\sigma_2^2|\boldsymbol{\alpha}_2, \ldots &\sim \text{I}\Gamma\left(\lambda + \frac{\text{rank}(\mathbf{G})}{2}, \psi + \frac{1}{2}\boldsymbol{\alpha}_2{}^\mathsf{T}\mathbf{G}\boldsymbol{\alpha}_2\right), \text{ and} \\
\sigma_\varepsilon^2|\boldsymbol{y}, \ldots &\sim \text{I}\Gamma\left(\frac{n}{2}, \frac{1}{2}\left(\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{U}\boldsymbol{\alpha}_1 - \mathbf{V}\boldsymbol{\alpha}_2\right)^\mathsf{T}\left(\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{U}\boldsymbol{\alpha}_1 - \mathbf{V}\boldsymbol{\alpha}_2\right)\right).
\end{aligned}$$

The remaining posterior distributions are similarly straightforward, but are less importance.

There is quite an extensive library of Bayesian methods for these classes of spatial statistics, including Markov Chain Monte Carlo (MCMC), i.e. Rue and Held (2005) and Gilks *et al.* (1995), and Integrated Nested Laplace Approximations (INLA), see Rue *et al.* (2009) and Martino and Chopin (2008).

These techniques are not readily applicable to RMVs though, as the conversion of the support of $\mathcal{D}$ to a lattice structure is far from straightforward. Also, a Bayesian formulation requires a great deal of computing time and power. For these reasons, this formulation was not explored.

4.3 Spatial Backfitting

The backfitting algorithm has been employed on a wide range of additive models, e.g. Furrer and Sain (2009), Buja *et al.* (1989), Breiman and Friedman (1985). The backfitting algorithm consists of iteratively estimating the fixed effects and the spatial terms until the estimates converge, i.e. the estimates no longer differ with each iteration.

The classical model for the backfitting algorithm is $\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta}+\boldsymbol{\alpha}+\boldsymbol{\varepsilon}$, where $\mathrm{Var}(\boldsymbol{\alpha}) = \boldsymbol{\Sigma}(\boldsymbol{\theta})$ and $\mathrm{Var}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}$. If the parameters $\boldsymbol{\theta}$ and σ are known, the backfitting algorithm then estimates the fixed effects using generalized least squares and the spatial effects using spatial smoothing:

$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}^\mathsf{T}\boldsymbol{\Sigma}_{\boldsymbol{y}}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^\mathsf{T}\boldsymbol{\Sigma}_{\boldsymbol{y}}^{-1}\boldsymbol{y} \quad \text{(generalized least-squares estimator)}, \tag{4.5}$$

$$\widehat{\boldsymbol{\alpha}} = \boldsymbol{\Sigma}(\boldsymbol{\theta})\boldsymbol{\Sigma}_{\boldsymbol{y}}^{-1}\left(\boldsymbol{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}\right) \quad \text{(spatial smoothing)}, \tag{4.6}$$

where $\boldsymbol{\Sigma}_{\boldsymbol{y}} = \mathrm{Var}(\boldsymbol{y}) = \boldsymbol{\Sigma}(\boldsymbol{\theta}) + \sigma^2\mathbf{I}$, the covariance matrix of the observations.

In practice, the parameters $\boldsymbol{\theta}$ and σ are unknown and they must be estimated in the iterative procedure.

Multiple Additive Components

Assume the RMVs can be decomposed by the additive decomposition

$$\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\alpha}_1 + \cdots + \boldsymbol{\alpha}_k + \boldsymbol{\varepsilon},$$

where each $\boldsymbol{\alpha}_i$ represents a different scale of variation. The backfitting algorithm for such a model is represented below.

Multiple Additive Spatial Terms Backfitting Algorithm

[0] Let $\widehat{\boldsymbol{\alpha}}_i^{(0)}$, for $i = 1, \ldots, k$, be an initial guess and put $j = 0$

[1] $j = j + 1$

[2] $\widehat{\boldsymbol{\beta}}^{(j)} = (\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}\mathbf{X}^\mathsf{T}(\boldsymbol{y} - \sum_{i=1}^{k}\widehat{\boldsymbol{\alpha}}_i^{(j-1)})$

[3] For $i = 1, \ldots, k$, estimate covariance parameters to get $\widehat{\sigma}^{2,(j)}$ and $\widehat{\boldsymbol{\Sigma}}_i^{(j)}$, then put $\widehat{\boldsymbol{\alpha}}_i^{(j)} = \widehat{\mathbf{W}}_i^{(j)}(\boldsymbol{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}^{(j)} - \sum_{\ell=1}^{i-1}\widehat{\boldsymbol{\alpha}}_\ell^{(j)} - \sum_{\ell=i+1}^{k}\widehat{\boldsymbol{\alpha}}_\ell^{(j-1)})$, $i = 1, \ldots, n$

[4] Repeat [2] to [4] until convergence.

In step [3] of the algorithm, $\widehat{\mathbf{W}}_i^{(j)} := \widehat{\boldsymbol{\Sigma}}_i^{(j)}(\widehat{\sigma}^{2,(j)}\mathbf{I} + \widehat{\boldsymbol{\Sigma}}_i^{(j)})^{-1}$. For convenience, we write $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{y},i}^{(j)} = \widehat{\sigma}^{2,(j)}\mathbf{I} + \widehat{\boldsymbol{\Sigma}}_i^{(j)}$.

Sequential Observations

The backfitting routine can also be used with rather complex covariance structures in the model. For example, each layer of the compaction process could be modeled as a new spatial process. Since the roller measures to a depth of several layers, the RMV recorded would thus be an observation of an additive process.

Let the first layer of material be modeled as

$$\boldsymbol{y}_1 = \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\alpha}_1 + \boldsymbol{\varepsilon}_1,$$

where $\mathbf{X}_1$ is full rank, $\boldsymbol{\alpha}_1$ is a zero mean spatial process independent of the zero mean, uncorrelated error process $\boldsymbol{\varepsilon}_1$. The second layer of the compaction site could then be modeled as

$$\boldsymbol{y}_2 = \mathbf{X}_2\boldsymbol{\beta}_2 + c\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2 + \boldsymbol{\varepsilon}_2,$$

where $\boldsymbol{\alpha}_2$ is the layer specific zero mean spatial process and c is a constant governing the "amount" of the first layer that is "seen" by the roller. Also, $\boldsymbol{\alpha}_2$ is independent of $\boldsymbol{\alpha}_1$ and $\boldsymbol{\varepsilon}_2$. Additional layers of compaction can be similarly modeled.

The parameters of such a model can be estimated using a sequential, spatial backfitting procedure for p total layers of compaction. The sequential, spatial backfitting procedure is an application of the backfitting routine in a sequential manner to each layer of the compaction:

Sequential Backfitting Algorithm

[$0^\dagger$] For $t = 1$ to p

[$1^\dagger$] Let $\widehat{\boldsymbol{\alpha}}_t^{(0)}$ be an initial guess and put $j = 0$

[$2^\dagger$] $j = j + 1$

[$3^\dagger$] $\widehat{\boldsymbol{\beta}}_t^{(j)} = (\mathbf{X}_t{}^\mathsf{T}\mathbf{X}_t)^{-1}\mathbf{X}_t{}^\mathsf{T}(\boldsymbol{y}_t - \sum_{i=1}^{t}\widehat{\boldsymbol{\alpha}}_i^{(j-1)})$

[$4^\dagger$] Estimate covariance parameters to get $\widehat{\sigma}_t^{2,(j)}$ and $\widehat{\boldsymbol{\Sigma}}_t^{(j)}$, then put
$\widehat{\boldsymbol{\alpha}}_t^{(j)} = \widehat{\mathbf{W}}_t^{(j)}(\boldsymbol{y}_t - \mathbf{X}_t\widehat{\boldsymbol{\beta}}_t^{(j)} - \sum_{\ell=1}^{t-1}\widehat{\boldsymbol{\alpha}}_\ell^{(j)})$

[$5^\dagger$] Repeat [$1^\dagger$] to [$5^\dagger$] until convergence.

In step [$4^\dagger$] of the algorithm,

$$\widehat{\mathbf{W}}_t^{(j)} = \widehat{\boldsymbol{\Sigma}}_t^{(j)}(\widehat{\sigma}_t^{2,(j)}\mathbf{I} + \widehat{\boldsymbol{\Sigma}}_t^{(j)})^{-1} = \widehat{\boldsymbol{\Sigma}}_t^{(j)}(\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{y},t}^{(j)})^{-1}.$$

For proofs of convergence, see Paper III.

Thesis Summary

This thesis consists of five papers, presented in chronological order. We first briefly summarize the contents of each paper.

Paper I

In Paper I, **Spatial Analysis of Modern Soil Compaction Roller Measurement Values** by Daniel K. Heersink and Reinhard Furrer, a model and estimation procedure for one layer of roller measurement values (RMVs) is proposed. Modern compaction rollers monitor soil properties by observing characteristics of the soil. A vibrating drum traverses the compaction site measuring soil stiffness and collecting GPS coordinates that are together termed RMVs. These RMVs can be modeled as a random spatial field and additively decomposed into any sensible combination of mean terms, spatial terms, spline terms, and ridge regression terms. The goal of this modeling is to implement intelligent compaction for quality control and quality assurance purposes. Proper modeling of such data (stationarity, anisotropy, ...) is then of paramount concern.

Each layer of the compaction site can be modeled by the n-vector $\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\alpha} + \boldsymbol{\gamma} + \boldsymbol{\varepsilon}$, where $\mathbf{X}\boldsymbol{\beta}$ is a low-order (linear) polynomial trend, $\boldsymbol{\alpha}$ is a mean term estimated using ridge regression or splines modeling the large-scale variation, $\boldsymbol{\gamma}$ is a (zero-mean Gaussian) spatial process modeling the small-scale variation, and $\boldsymbol{\varepsilon}$ is the noise. Here, $\mathbf{X}$ is the $(n \times p)$ design matrix with rank p. There are many general approaches to working with such an additive mixed model, including a backfitting procedure for maximum likelihood estimation and generalized cross validation. Due to computational complexity of maximum likelihood estimation a backfitting procedure, Furrer and Sain (2009), was extended to the more general models used here and employed in the estimation. The extended backfitting procedure has been shown to converge and the iterative least squares estimates have been shown to converge to the generalized least squares estimate.

A simulation study has been conducted to analyze estimates of this general model using a penalized likelihood and generalized cross validation (GCV) approach as well. Results of the cross validation study using a spline structure indicate there are some random fields that can be generated that do not have a minimum GCV.

Note: $\mathbf{H}$ is defined as symmetric, semi-positive definite. In the case of additive ridge regression terms, $\mathbf{H}$ is in fact positive definite. Thus for these terms, $\mathbf{H}^{-1}$ exists. In the case of spline terms, $\mathbf{H}$ is semi-positive definite and corresponds to the matrix of basis functions. In this case, $\mathbf{H}^{-1}$ is understood to mean the Moore–Penrose pseudoinverse of the matrix and the resulting probability density function $\boldsymbol{\alpha}|\mathbf{H}, \sigma_\alpha^2 \sim \mathcal{N}(\mathbf{0}, \sigma_\alpha^2 \mathbf{H}^{-1})$ does not exist. Refer to Section 4.2.

The main contribution of the paper is the development of a stochastic model for one layer of RMVs. The groundwork is provided to further explore the modeling and estimation of compaction data with several potential avenues of research outlined.

Paper II

In Paper II, **On Moore–Penrose Inverses of quasi-Kronecker Structured Matrices** by Daniel K. Heersink and Reinhard Furrer, the rank one update to the Moore–Penrose pseudoinverse is extended to larger rank updates. The Moore–Penrose pseudoinverse and generalized inverse of $\mathbf{A} + \mathbf{X}_1\mathbf{X}_2^*$, where $\mathbf{A}, \mathbf{X}_1, \mathbf{X}_2$ are complex matrices are given under various assumptions.

We use the result to derive the Moore–Penrose pseudoinverse and inverse for $\mathrm{bdiag}(\mathbf{A}_k) + \boldsymbol{u}\boldsymbol{v}^* \otimes \mathbf{E}$ with p complex matrices $\mathbf{A}_k$ of dimension $n \times m$, two complex p-vectors $\boldsymbol{u}$ and v and a complex matrix $\mathbf{E}$ of dimension $n \times m$. Such block structured matrices occur in hierarchical modeling of multivariate spatial and spatio-temporal Gaussian processes. For the latter, expressions of the determinant and of conditional variances are provided.

The main contribution of the paper is the extension of the rank one update to the Moore–Penrose pseudoinverse to several cases of higher rank updates. These extensions are applied to the quasi-Kronecker structured matrix that appears in hierarchical modeling of multivariate spatial and spatio-temporal processes.

Paper III

In Paper III, **Sequential Spatial Analysis of Large Datasets with Applications to Modern Earthwork Compaction Roller Measurement Values** by Daniel K. Heersink and Reinhard Furrer, a sequential, spatial mixed-effects model is proposed and a sequential, spatial backfitting algorithm is developed for its estimation and applied to roller measurement values from a test site in Minnesota, USA.

In the context of road construction, modern earthwork compaction rollers equipped with sensors collect a virtually continuous flow of soil property measurements. This sequential, spatial data can be utilized to improve the quality control of the compaction process through the introduction of intelligent compaction.

These roller measurement values are observed indirectly through non-linear measurement operators, non-stationary, inherently multivariate with complex correlation structures, and collected in huge quantities. The problem of modeling and estimation in a spatially correlated setting with large amounts of data is well known and many approaches can be found in the literature.

Due to the complexity of the correlation structures, a simple model is often utilized in these approaches. Very few studies have been completed investigating sequential, spatially correlated data outside of a point process framework. We propose a sequential, spatial mixed-effects model and develop a sequential, spatial backfitting algorithm to estimate fixed effects and several independent, spatially correlated processes. This new algorithm is demonstrated in a simulation study and applied to earthwork compaction data.

Statistical models previously proposed in the literature do not adequately address such data. Toward that end, a sequential, spatial mixed-effects model is proposed and a backfitting algorithm is developed to estimate fixed effects and several independent, spatially correlated processes. The quasi-Kronecker structure of Heersink and Furrer (2012) can also be implemented to aid

in computation time. This new algorithm is demonstrated in a simulation study and applied to earthwork compaction data.

The main contribution of the paper is the development of the sequential, spatial mixed-effects model and the sequential, spatial backfitting routine for estimation of such models. The data analyzed in this paper was provided by Dr. Mike Mooney and his team at Colorado School of Mines.

Paper IV

In Paper IV, **Intelligent Compaction and Quality Assurance of Roller Measurement Values utilizing Backfitting and Multiresolution Scale Space Analysis** by Daniel K. Heersink, Reinhard Furrer, and Mike A. Mooney, the sequential, spatial mixed-effects model and Sequential Backfitting Algorithm developed in Paper III are applied to a dataset consisting of three layers of compaction material to estimate the true field. These estimates are then treated as images and a scale space multiresolution analysis is performed to identify credible regions of hard and soft spots. This implementation is designed for improvement of intelligent compaction and quality control of the compaction process.

The main contribution of the paper is the utilization of the backfitting algorithm, coupled with multiresolution analysis as a means of quality control and intelligent compaction for modern earthwork compaction. This model formulation, estimation procedure, and image analysis regime utilizes the uncertainty in RMVs to provide truly intelligent compaction capabilities. The data used in this paper was provided by Dr. Mike Mooney and his team at Colorado School of Mines and the image analysis programming was provided by Dr. Lasse Holmström and his team at University of Oulu.

Paper V

In Paper V, **Spatial Backfitting of Roller Measurement Values from a Florida Test Bed** by Daniel K. Heersink, Reinhard Furrer, and Mike A. Mooney, a large, atypically compacted test bed dataset from Florida is analyzed. Preliminary empirical semivariograms and detrending are performed to investigate anisotropy concerns. To investigate the importance of driving direction the Sequential Backfitting Algorithm from Paper III is implemented.

The main contribution of the paper is an application of the backfitting algorithm to a large test bed for the analysis of driving direction dependence on RMVs. The data used in this paper was provided by Dr. Mike Mooney and his team at Colorado School of Mines.

Bibliography

Adam, D. and Kopf, F. (2004). Operational devices for compaction optimization and quality control. In *Geotechnics in Pavement and Railway Design and Construction: Proceedings of the International Seminar on Geotechnics and Railway Design and Construction, Athens, Greece, 16 - 17 December 2004*, 97–106. IOS Press.

Anderegg, R. and Kaufmann, K. (2004). Intelligent compaction with vibratory rollers: Feedback control systems in automatic compaction and compaction control. *Transportation Research Record: Journal of the Transportation Research Board*, **1868**, 124–134.

Breiman, L. and Friedman, J. H. (1985). Estimating optimal transformations for multiple regression and correlation. *Journal of the American Statistical Association*, **80**, 580–598.

Buja, A., Hastie, T., and Tibshirani, R. (1989). Linear smoothers and additive models. *The Annals of Statistics*, **17**, 453–510.

Chilès, J. P. and Delfiner, P. (1999). *Geostatistics Modeling Spatial Uncertainty*. Wiley.

Cressie, N. (1993). *Statistics for spatial data*. Wiley series in probability and mathematical statistics: Applied probability and statistics. J. Wiley.

Cressie, N. and Hawkins, D. M. (1980). Robust estimation of the variogram. *Journal of the International Association of Mathematical Geology*, **1**, 115–125.

Cressie, N. and Wikle, C. K. (2011). *Statistics for Spatio-Temporal Data (Wiley Series in Probability and Statistics)*. Wiley.

Eilers, P. H. C. (1991). Penalized regression in action: Estimating pollution roses from daily averages. *Environmetrics*, **2**, 25–47.

Eilers, P. H. C. (2003). A perfect smoother. *Analytical Chemistry*, **75**, 3631–3636.

Eilers, P. H. C. and Marx, B. D. (2004). Splines, knots and penalties. Technical report.

Eilers, P. H. C., Rijnmond, D. M., and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, **11**, 89–121.

Facas, N. W. (2009). Variogram properties and anisotropy in the spatial distribution of roller-measured soil stiffness. Master's thesis, Colorado School of Mines.

Facas, N. W. and Mooney, M. A. (2011). Characterizing the precision uncertainty in vibratory roller measurement values. *Journal of Testing and Evaluation*, **40**, 1–9.

Furrer, R. and Sain, S. R. (2009). Spatial model fitting for large datasets with applications to climate and microarray problems. *Statistics and Computing*, **19**, 113–128.

Gilks, W., Richardson, S., and Spiegelhalter, D., editors (1995). *Markov Chain Monte Carlo in Practice (Chapman & Hall/CRC Interdisciplinary Statistics)*. Chapman and Hall/CRC, 1st edition.

Gneiting, T. and Guttorp, P. (2010). Continuous parameter stochastic process theory. In Gelfand, A., Diggle, P., Fuentes, M., and Guttorp, P., editors, *Handbook of Spatial Statistics*, Chapman & Hall/CRC Handbooks of Modern Statistical Methods, chapter 2, 17–28. CRC Press.

Heersink, D. K. and Furrer, R. (2011). Spatial analysis of modern soil compaction roller measurement values. *Procedia Environmental Sciences*, **7**, 8 – 13.

Heersink, D. K. and Furrer, R. (2012). On Moore–Penrose inverses of quasi-Kronecker structured matrices. *Linear Algebra and its Applications*, **436**, 561–570.

Heersink, D. K. and Furrer, R. (2013). Sequential spatial analysis of large datasets with applications to modern earthwork compaction roller measurement values. Submitted to *Spatial Statistics*.

Heersink, D. K., Furrer, R., and Mooney, M. A. (2013a). Intelligent compaction and quality assurance of roller measurement values utilizing backfitting and multiresolution scale space analysis. arXiv:1302.4631.

Heersink, D. K., Furrer, R., and Mooney, M. A. (2013b). Spatial backfitting of roller measurement values from a Florida test bed. arXiv:1302.4659.

Krige, D. G. (1951). A Statistical Approach to Some Basic Mine Valuation Problems on the Witwatersrand. *Journal of the Chemical, Metallurgical and Mining Society of South Africa*, **52**, 119–139.

Kröber, W., Floss, R., and Wallrath, W. (2001). Dynamic soil stiffness as quality criterion for soil compaction. In Correia, A. G. and Brandl, H., editors, *Geotechnics for Roads, Rail Tracks, and Earth Structures*, 189–199. Taylor & Francis.

Martino, S. and Chopin, N. (2008). Implementing approximate bayesian inference for latent gaussian models using integrated nested laplace approximations: A manual for the inla-program. Technical report, Norwegian University of Science and Technology.

Marx, B. D., Eilers, P. H., and Li, B. (2011). Multidimensional single-index signal regression. *Chemometrics and Intelligent Laboratory Systems*, **109**, 120 – 130.

Marx, B. D. and Eilers, P. H. C. (1998). Direct generalized additive modeling with penalized likelihood. *Computational Statistics & Data Analysis*, **28**, 193–209.

Matérn, B. (1986). *Spatial variation*. Lecture notes in statistics. Springer-Verlag.

Matheron, G. (1962). *Traité de Géostatistique Appliquée*. Tome 1, Editions Technip, Paris.

Matheron, G. (1963). Principles of geostatistics. *Economic Geology*, **58**, 1246–1266.

Mooney, M. and Adam, D. (2007). Vibratory roller integrated measurement of earthwork compaction: An overview. volume 307, 80–80. ASCE.

Mooney, M. A. and Rinehart, R. V. (2007). Field monitoring of roller vibration during compaction of subgrade soil. *Journal of Geotechnical and Geoenvironmental Engineering*, **133**, 257–265.

Mooney, M. A. and Rinehart, R. V. (2009). In situ soil response to vibratory loading and its relationship to roller-measured soil stiffness. *Journal of Geotechnical and Geoenvironmental Engineering*, **135**, 1022–1031.

Mooney, M. A., Rinehart, R. V., White, D. J., Vennapusa, P. K., Facas, N. W., and Musimbi, O. M. (2010). Intelligent soil compaction systems: NCHRP project 21-09 final report.

Perperoglou, A. and Eilers, P. H. C. (2010). Penalized regression with individual deviance effects. *Computational Statistics*, **25**, 341–361.

Petersen, D., Erickson, M., Roberson, R., and Siekmeier, J. (2007). Intelligent soil compaction: geostatistical data analysis and construction specifications. In *Transportation Research Board 86th Annual Meeting*.

Puppala, A. (2008). Estimating stiffness of subgrade and unbound materials for pavement design: NCHRP project 20-5.

Rinehart, R. and Mooney, M. (2008). Measurement depth of vibratory roller-measured soil stiffness. *Géotechnique*, **59**, 609–619.

Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory And Applications*. Monographs on Statistics and Applied Probability. Chapman & Hall/CRC.

Rue, H., Martino, S., and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **71**, 319–392.

Schabenberger, O. and Gotway, C. A. (2005). *Statistical methods for spatial data analysis*. Texts in Statistical Science Series. Chapman & Hall/CRC, Boca Raton, FL.

Scherocman, J., Rakowski, S., and Uchiyama, K. (2007). Intelligent compaction, does it exist? In *Proceedings of the 52 Annual Conference of the Canadian Technical Asphalt Association*. Polyscience Publications Inc.

Senseney, C. and Mooney, M. (2010). Characterization of a two-layer soil system using a light weight deflectometer with radial sensors. *Journal of the Transportation Research Board*, **1**, 21–28.

Stein, M. L. (1999). *Interpolation of Spatial Data*. Springer-Verlag, New York.

Thurner, H. and Sandström, A. (1980). A new device for instant compaction control. In *Proceedings of International Conference on Compaction*, volume II, 611–614.

van Susante, P. J. and Mooney, M. A. (2008). Capturing nonlinear vibratory roller compactor behavior through lumped parameter modeling. *Journal of Engineering Mechanics*, **134**, 684–693.

Wahba, G. (1990). *Spline Models for Observational Data*. CBMS-NSF Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Mathematics.

White, D. J. and Thompson, M. J. (2008). Relationships between in situ and roller-integrated compaction measurements for granular soils. *Journal of Geotechnical and Geoenvironmental Engineering*, **134**, 1763–1770.

Wood, S. (2006). *Generalized Additive Models: An Introduction with R*. Texts in Statistical Science. Chapman & Hall/CRC.

Yoo, T. and Selig, E. (1979). Dynamics of vibratory-roller compaction. *Journal of the Geotechnical Engineering Division*, **105**, 1211–1231.

Zimmerman, D. L. (1993). Another look at anisotropy in geostatistics. *Mathematical Geology*, **25**, 453–470.

Paper I

Spatial Analysis of Modern Soil Compaction Roller Measurement Values

Daniel K. Heersink & Reinhard Furrer

Conference Paper published in *Procedia Environmental Sciences* (2011), pp. 8–13.

Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Environmental Sciences 7 (2011) 8–13

Procedia Environmental Sciences

1st Spatial Statistics Conference 2011: Mapping Global Change

Spatial analysis of modern soil compaction roller measurement values

Daniel K. Heersink[a*], Reinhard Furrer[a]

[a]*University of Zurich, Winterthurerstrasse 190, 8049 Zurich, Switzerland*

Abstract

Modern compaction rollers monitor soil properties by observing vibrational characteristics of the soil. A vibrating drum traverses the compaction site measuring soil stiffness and collecting GPS coordinates that are together termed roller measurement values (RMVs). These RMVs can be modeled as a random spatial field and additively decomposed into any sensible combination of mean terms, spatial terms, spline terms, and ridge regression terms. The goal of this modeling is to implement intelligent compaction for quality control and quality assurance purposes. Proper modeling of such data (stationarity, anisotropy, . . .) is then of paramount concern.

Each layer of the compaction site can be modeled by the n-vector $\mathbf{y} = \mathbf{X\beta} + \boldsymbol{\alpha} + \boldsymbol{\gamma} + \boldsymbol{\varepsilon}$, where $\mathbf{X\beta}$ is a low-order (linear) polynomial trend, $\boldsymbol{\alpha}$ is a mean term estimated using ridge regression or splines modeling the large-scale variation, $\boldsymbol{\gamma}$ is a (zero-mean Gaussian) spatial process modeling the small-scale variation, and $\boldsymbol{\varepsilon}$ is the noise. Here, $\mathbf{X}$ is the $(n \times p)$ design matrix with rank p. There are many general approaches to working with such an additive mixed model, including a backfitting procedure for maximum-likelihood estimation and generalized cross-validation.

Due to computational complexity of maximum-likelihood estimation a backfitting procedure, Furrer and Sain (2009) [1], was extended to the more general models used here and employed in the estimation. The extended backfitting procedure has been shown to converge and the iterative least-squares estimates have been shown to converge to the generalized least-squares estimate.

A simulation study has been conducted to analyze estimates of this general model using a penalized likelihood and generalized cross-validation (GCV) approach as well. Results of the cross-validation study using a spline structure indicate there are some random fields that can be generated that do not have a minimum GCV.

Keywords: generalized cross-validation; backfitting; stationarity; anisotropy

* Corresponding author. Tel.: +41-(0)44-635-5870; fax: +41-(0)44-635-5706.
E-mail address: daniel.heersink@math.uzh.ch.

1878-0296 © 2011 Published by Elsevier Ltd. Selection and peer-review under responsibility of Spatial Statistics 2011
doi:10.1016/j.proenv.2011.07.003

1. Introduction

Modern compaction rollers for road construction are outfitted with sensors that record and output a measure of the underlying material stiffness. A representative roller manufactured by Bomag can be found in Figure 1. This output is a relative measure that can aid in quality assurance and quality control (QA/QC) of the compaction area. The sensor output, coupled with GPS coordinates, is termed the roller measurement value (RMV). These measurements are a complicated, weighted measure of an approximately $2m^3$ bulb of underlying material. New RMVs are reported every 2-5cm in the direction of driving (x-direction) leading to very dense data. Conversely, the RMVs are very sparse in the transverse direction (y-direction) as they are usually 1-2m apart. Typical construction techniques involve compaction of several successive layers of material 15-30cm thick. Figure 2 is an example of the output of such a sensor equipped roller. Note that data is reported as a point but displayed as a box to better represent its areal nature.

Figure 1: Bomag compaction roller on a test site

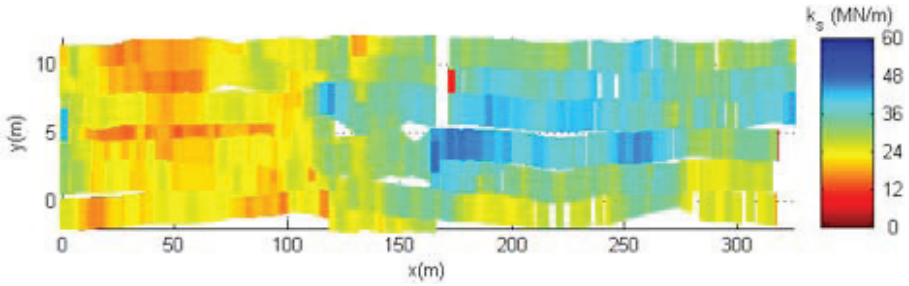

Figure 2: RMV data from a test field

We model RMV data as an additive mixed model (AMM). Each layer of the compaction site is modeled as an additive decomposition of a polynomial trend, a ridge regression or spline term, a (zero-mean Gaussian) spatial process, and white noise. The "correlation range" of the ridge regression/spline term is much larger than that of the spatial process (i.e., smoother fields compared to the spatial process) as to maintain identifiability of the two terms. The highly non-stationary and anisotropic nature of the data requires the use of both the spatial process and the ridge regression/spline term.

Due to the very large amount of data, traditional estimation methods such as maximum likelihood are too cumbersome so we utilize a backfitting algorithm to estimate the coefficients and variance parameters

in the model. Generalized cross-validation (GCV) can also be employed as an alternate estimation procedure.

2. Nature of RMVs and uncertainty in RMVs

Every roller manufacturer produces a unique measure of soil stiffness. This stiffness measure, coupled with GPS coordinates, defines the RMV. RMVs are calculated from frequency and excitation force amplitude information gathered from sensors within the vibrating and rotating smooth drum. RMVs are directly proportional to the stiffness of the underlying soil [2]. The Bomag vibration modulus (E_{vib}) and Ammann/CASE soil stiffness (k_s) RMVs directly represent soil stiffness by calculating the contact force and displacement during vibration [3], [4]. The Dynapac and Caterpillar compaction meter value (CMV) and Sakai continuous compaction value (CCV) RMVs indirectly represent soil stiffness by giving a measure of the nonlinearity associated with the loss of contact during vibration [5], [6], [7]. Many studies and models have verified that these RMVs represent soil stiffness [8], [7], [9], [10].

These RMVs reflect the aggregate stiffness of the underlying material to approximately 1m [11]. Due to this complication of multiple layers of material, the RMVs are often correlated to industry standard spot tests [12]. This correlation between RMVs and spot test measurements is somewhat suspect given the different measurement depths of the two devices (1m vs. 0.2-0.3m).

The bias of RMVs cannot be determined given the complex nature of the compaction process. Therefore, RMVs are a relative measure of soil stiffness [13]. This is not as hindering as it looks as many industry standard spot tests are likewise relative measures of soil stiffness, e.g., falling weight deflectometer, light weight deflectometer, and soil stiffness gage [14]. Therefore, the uncertainty needing quantification is that termed precision, the repeatability of measurements.

There are two values of uncertainty in RMVs. The uncertainty in a single RMV is σ_1 and the uncertainty in the difference between two different RMVs at the same spatial location is σ_2. Assuming that the error is Gaussian with variance σ_1^2, the two uncertainties are related by $\sigma_2^2 = 2\sigma_1^2$.

3. Statistical Model

The RMVs are modeled as: $\mathbf{y} = \mathbf{X\beta} + \boldsymbol{\alpha} + \boldsymbol{\gamma} + \boldsymbol{\varepsilon}$, where $\mathbf{X\beta}$ is a (low-order) polynomial trend, $\boldsymbol{\alpha}$ is a ridge regression term or spline term, $\boldsymbol{\gamma}$ is a (zero-mean Gaussian) spatial process, and $\boldsymbol{\varepsilon}$ is white noise. Here, $\mathbf{X}$ has full rank. Now, we use a hierarchical Bayesian model framework to model $\mathbf{y}$ conditionally as $\mathbf{y}|\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \sigma_\varepsilon^2 \sim \mathcal{N}(\mathbf{X\beta} + \boldsymbol{\alpha} + \boldsymbol{\gamma}, \sigma_\varepsilon^2\mathbf{I})$. We model $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$: $\boldsymbol{\alpha}|\mathbf{H}, \sigma_\alpha^2 \sim \mathcal{N}(\mathbf{0}, \sigma_\alpha^2\mathbf{H}^{-1})$, $\boldsymbol{\gamma}|\mathbf{G}, \sigma_\gamma^2 \sim \mathcal{N}(\mathbf{0}, \sigma_\gamma^2\mathbf{G}^{-1})$, where $\mathbf{H}$ is symmetric positive semi-definite, $\mathbf{G}$ is symmetric positive definite, and $\boldsymbol{\alpha}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\varepsilon}$ are independent. The variance parameters are then given (improper) inverse-gamma prior distributions: $\sigma_\alpha^2 \sim \mathcal{IG}(\eta, \nu)$, $\sigma_\gamma^2 \sim \mathcal{IG}(\lambda, \psi)$, $\sigma_\varepsilon^2 \sim 1/\sigma_\varepsilon$. These prior distributions lead to straightforward, analytic posterior distributions: $\sigma_\alpha^2|\boldsymbol{\alpha}, \mathbf{H} \sim \mathcal{IG}(\eta + \text{rank}(\mathbf{H})/2, \nu + 1/2\boldsymbol{\alpha}^T\mathbf{H}\boldsymbol{\alpha})$, $\sigma_\gamma^2|\boldsymbol{\gamma}, \mathbf{G} \sim \mathcal{IG}(\lambda + n/2, \psi + 1/2\boldsymbol{\gamma}^T\mathbf{G}\boldsymbol{\gamma})$, $\sigma_\varepsilon^2|\mathbf{y}, \ldots \sim \mathcal{IG}(n/2, 1/2(\mathbf{y} - \mathbf{X\beta} - \boldsymbol{\alpha} - \boldsymbol{\gamma})^T(\mathbf{y} - \mathbf{X\beta} - \boldsymbol{\alpha} - \boldsymbol{\gamma}))$. Estimation of model parameters and components in this model is quite difficult due to the large amount of data and the general lack of sparsity in the covariance structure. Two estimation methods have thus been developed for complex models of this nature.

3.1. Generalized Cross Validation (GCV)

The first estimation procedure is a penalized likelihood approach where the optimal smoothing parameters are chosen using generalized cross validation (GCV). Let $\lambda_\alpha, \lambda_\gamma > 0$ be the smoothing parameters for $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ respectively. The goal of the estimation is then to minimize $||\mathbf{y} - \mathbf{X\beta} - \boldsymbol{\alpha} - \boldsymbol{\gamma}||^2 +$

$\lambda_\alpha \boldsymbol{\alpha}^T\mathbf{H}\boldsymbol{\alpha} + \lambda_\gamma \boldsymbol{\gamma}^T\mathbf{G}\boldsymbol{\gamma}$. Let $\mathbf{W} = [\mathbf{X}\ \mathbf{I}\ \mathbf{I}]$, $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma})^T$, and $\boldsymbol{\Omega} = \mathrm{diag}(\mathbf{0}, \mathbf{H}, \lambda_\alpha/\lambda_\gamma\mathbf{G})$. We can now reformulate the problem as the well known problem of minimizing $\|\mathbf{y} - \mathbf{W}\boldsymbol{\theta}\|^2 + \lambda_\alpha\boldsymbol{\theta}^T\boldsymbol{\Omega}\boldsymbol{\theta}$.

The hat matrix is thus $\mathbf{A}(\lambda_\alpha,\lambda_\gamma) = \mathbf{W}(\mathbf{W}^T\mathbf{W} + \lambda_\alpha\boldsymbol{\Omega})^{-1}\mathbf{W}^T$ and the trace of $\mathbf{A}(\lambda_\alpha,\lambda_\gamma)$ gives an estimate of the degrees of freedom, or effective number of parameters, in the model.

The minimization problem is thus the simultaneous solution of:

- $\mathbf{X}^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \boldsymbol{\alpha} - \boldsymbol{\gamma}) = \mathbf{0}$,
- $\boldsymbol{\alpha} = (\lambda_\alpha + 1)^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \boldsymbol{\gamma})$, and
- $\boldsymbol{\gamma} = (\lambda_\gamma + 1)^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \boldsymbol{\alpha})$.

Thus,

- $\mathbf{b} = (\mathbf{X}^T\mathbf{R}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{R}^{-1}\mathbf{y}$,
- $\mathbf{g} = \lambda_\alpha(\lambda_\gamma\mathbf{M} + \lambda_\alpha\mathbf{I})^{-1}(\mathbf{y} - \mathbf{X}\mathbf{b})$, and
- $\mathbf{a} = \mathbf{M}^{-1}(\mathbf{y} - \mathbf{X}\mathbf{b} - \mathbf{g})$.

Here, $\mathbf{R}^{-1} = \lambda_\alpha\lambda_\gamma/(\lambda_\alpha\lambda_\gamma + \lambda_\alpha + \lambda_\gamma)\mathbf{I}$ and $\mathbf{M} = (\lambda_\alpha + 1)\mathbf{I}$.

Using a Bayesian approach to parameter selection, we could also get posterior estimates of the smoothing parameters. If $\boldsymbol{\alpha} \sim \mathcal{N}(\mathbf{0}, \sigma_\alpha^2\mathbf{H}^{-1})$ and $\boldsymbol{\gamma} \sim \mathcal{N}(\mathbf{0}, \sigma_\gamma^2\mathbf{G}^{-1})$, independent of $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \sigma_\varepsilon^2\mathbf{I})$, then it is straightforward to show that $\lambda_\alpha = \sigma_\varepsilon^2/\sigma_\alpha^2$ and $\lambda_\gamma = \sigma_\varepsilon^2/\sigma_\gamma^2$.

3.2. Backfitting Algorithm

A backfitting algorithm for spatial data was developed in [1]. The model used there is $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\gamma} + \boldsymbol{\varepsilon}$, where $\boldsymbol{\gamma}$ is a zero-mean spatial process. This has been extended to two additive modeling terms $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ that can take the form of zero-mean spatial processes, ridge regression terms, or spline terms. The extended algorithm is thus:

1. Let $\mathbf{a}^{(0)}$ and $\mathbf{g}^{(0)}$ be an initial guess for $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ respectively and put $j = 0$
2. $j = j + 1$
3. $\mathbf{b}^{(j)} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\mathbf{y} - \mathbf{a}^{(j-1)} - \mathbf{g}^{(j-1)})$
4. Estimate covariance parameters $s_\alpha^{(j)}$, $s_\gamma^{(j)}$, and $s_\varepsilon^{(j)}$ of the model, then put $\mathbf{a}^{(j)} = \mathbf{W}_\alpha^{(j)}(\mathbf{y} - \mathbf{X}\mathbf{b}^{(j)} - \mathbf{g}^{(j-1)})$ and $\mathbf{g}^{(j)} = \mathbf{W}_\gamma^{(j)}(\mathbf{y} - \mathbf{X}\mathbf{b}^{(j)} - \mathbf{a}^{(j)})$
5. Repeat 2 to 4 until convergence.

In step 4 of the algorithm, $\mathbf{W}_\alpha^{(j)} = (s_\alpha^{(j)})^2\mathbf{H}^{-1}((s_\varepsilon^{(j)})^2\mathbf{I} + (s_\alpha^{(j)})^2\mathbf{H}^{-1})^{-1}$ and $\mathbf{W}_\gamma^{(j)} = (s_\gamma^{(j)})^2\mathbf{G}^{-1}((s_\varepsilon^{(j)})^2\mathbf{I} + (s_\gamma^{(j)})^2\mathbf{G}^{-1})^{-1}$. This very general model setup allows for a wide range of model descriptions of the data. Taking advantage of the matrix identity $\mathbf{I} - \mathbf{A}(\mathbf{A} + \lambda\mathbf{I})^{-1} = \lambda(\mathbf{A} + \lambda\mathbf{I})^{-1}$, where $\mathbf{A}$ is semi-positive definite and $\lambda > 0$ allows us to extend the proof found in [1]. The basic concept behind the proof is to show equivalence between the iterative ordinary least-squares estimates of $\boldsymbol{\beta}$ with the generalized least-squares estimate.

By rewriting the ridge regression estimate as $\mathbf{a} = (\mathbf{I} + \lambda\mathbf{H})^{-1}(\mathbf{y} - \mathbf{X}\mathbf{b} - \mathbf{g}) = \mathbf{H}^{-1}(\lambda\mathbf{I} + \mathbf{H}^{-1})^{-1}(\mathbf{y} - \mathbf{X}\mathbf{b} - \mathbf{g})$, the estimate takes the same form as that of the spatial term. Thin-plate-spline (TPS) terms can also be estimated with a similar form as the spatial terms by letting the $\mathbf{H}$ be the generalized covariance matrix for the radial smoother. As all three of these formulations are mathematically equivalent, proof of equivalence needs only to be shown for one of the formulations and is a straightforward calculation.

It still remains to show the algorithm converges. Let $\mathbf{v}^{(j)} = (\mathrm{hat}(\mathbf{y})^{(j)}, \mathbf{a}^{(j)}, \mathbf{g}^{(j)})^T$ and $\mathbf{w} = (\mathbf{y}, \ldots, \mathbf{y})^T$., where $\mathrm{hat}(\mathbf{y})^{(j)}$, $\mathbf{a}^{(j)}$, $\mathbf{g}^{(j)}$ are smooth estimates of $\mathbf{y}$, $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ in iteration j. Define $\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}$. Then $\mathrm{hat}(\mathbf{y})^{(j)} = \mathbf{H}(\mathbf{y} - \mathbf{a}^{(j-1)} - \mathbf{g}^{(j-1)})$, $\mathbf{a}^{(j)} = \mathbf{W}_\alpha\mathbf{y} - \mathbf{W}_\alpha\mathrm{hat}(\mathbf{y})^{(j)} - \mathbf{W}_\alpha\mathbf{g}^{(j-1)}$, and $\mathbf{g}^{(j)} = \mathbf{W}_\gamma\mathbf{y} - \mathbf{W}_\gamma\mathrm{hat}(\mathbf{y})^{(j)} - \mathbf{W}_\gamma\mathbf{a}^{(j)}$. Thus we rewrite the problem as $\mathbf{v}^{(j)} = \boldsymbol{\Sigma}_{l=0\ to\ j-1}(-\mathbf{B})^l\mathbf{A}\mathbf{w}$, which converges if $(\mathbf{I} + \mathbf{B})^{-1}$ exists. The matrix $\mathbf{I} + \mathbf{B}$ can be rewritten in a quasi-kronecker structure. Thus, using several matrix algebra techniques, the inverse

exists and the algorithm converges. Simulation studies of the rate of convergence of the algorithm are yet to be completed.

4. Results

4.1. Simulation Study

Data was simulated on a rectangular grid $\boldsymbol{x}$ from the model $\mathbf{y} = \mathbf{X\beta} + \boldsymbol{\alpha} + \boldsymbol{\gamma} + \boldsymbol{\varepsilon}$, where $\mathbf{X} = (\mathbf{1}, \boldsymbol{x})$, $\boldsymbol{\beta} = (1, 1, 1)^T$, $\boldsymbol{\gamma}$ is a zero mean exponential random field, $\mathrm{Cov}(\boldsymbol{\alpha}) = \|\boldsymbol{x}\|^2 \log(\|\boldsymbol{x}\|)$, and $\boldsymbol{\varepsilon}$ is white noise. Parameter and variance estimates were then calculated using the penalized likelihood approach from above. Due to the large dataset, a standard optimization method is not feasible so a grid search on a grid of λ_α and λ_γ values is employed instead. The minimum on the grid, $\boldsymbol{\lambda}_{\min}$, was recorded for many realizations of the simulated data. The simulation was repeated for several different values of the signal to noise ratio for both $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$.

The minimum smoothing parameters tend to cluster at their true (theoretical) values. There is, however, a distinct clustering of values for some of the simulated fields at the boundary of the $\boldsymbol{\lambda}$ grid. As the dimension of the grid increases, these clustering points stay on the boundary. This may be due to numerical precision problems as the GCV function never reaches a minimum.

To investigate this boundary issue, data was simulated on the same grid after removing one of the three modeling elements from the model. The elimination of $\mathbf{X\beta}$ from the model does not eliminate the boundary clustering. Thus the mean term does not affect the GCV calculation or the boundary issues associated with the minimization.

Elimination of $\boldsymbol{\alpha}$ from the model successfully eliminates the boundary issues. This is an expected result as the model has been simplified to a constant mean term, a random field with exponential covariance function and a white noise. Elimination of $\boldsymbol{\gamma}$ from the model results in the same successful elimination of boundary issues. Both of these simplifications lead to well known models and the results are consistent with theory.

The grid was then collapsed to one dimension and the λ_γ value held constant at the true value. The resultant marginal GCV function for λ_α still displayed evidence of the boundary issues when minimizing in two dimensions. Calculating the marginal GCV while holding λ_α constant at its true value still produces boundary issues when minimizing the λ_γ values. These issues are independent of grid size and dimensionality.

However, one has to note that the estimation difficulty in the simulated datasets is not worrisome for practical aspects. For real observational RMVs we always found sensible values for the smoothing parameters.

4.2. Discussion

There are some random fields that can be generated that do not have a minimum GCV. These fields have not yet been thoroughly investigated to determine their properties. It is not yet known whether the boundary issues are a numerical accuracy issue or the data does not sufficiently discriminate the “range” parameters of both components.

To further investigate this latter problem, the spline part of the model was replaced with a Gaussian field with a Matérn covariance function and a range comparable to that of the spline structure. The smoothing parameter of this field, ν was set to be 1.5. Preliminary results of this investigation indicate the boundary issues are not necessarily related to the spline construction as there are still fields generated

with no minimum GCV. There are fewer of these fields generated though, meaning fewer λ_{min} found on the boundary.

Beyond the boundary problems, most values of λ_{min} are distributed around the true value. The scale of the variance in either the α or the γ direction is relative to the size of the true value of the parameter. For small values of the true parameter, the variance is smaller and vice versa.

5. Outlook and further research

Proper modeling of the multiple layering aspect of construction is still in progress. One approach is a multivariate spatial field that has a spatio-temporal aspect due to the consecutive compaction of soil layers. Another approach would be to use Gaussian Markov random fields (GMRFs) instead of spatial fields. A three dimensional lattice structure of the construction site could then be built. Spatial dependencies could then be defined on this lattice. The problem with this technique is the anisotropic nature of the process. Proper modeling of anisotropy is not completely understood using GMRFs.

Acknowledgements

This research was partially funded by the Swiss National Science Foundation Project 200021-129782.

References

[1] R. Furrer and S.R. Sain, Spatial model fitting for large datasets with applications to climate and microarray problems. *Statistics and Computing*, **19**, 113-128.

[2] N.W. Facas and M.A. Mooney, Characterizing the uncertainty in vibratory roller measurement values. *In review*.

[3] W. Kröber, R. Floss and W. Wallrath, Dynamic soil stiffness as quality criterion for soil compaction. In Correia, A. G. and Brandl, H., editors, *Geotechnics for Roads, Rail Tracks, and Earth Structures*, 189-199. Taylor & Francis.

[4] M.A. Mooney and R.V. Rinehart, Field monitoring of roller vibration during compaction of subgrade soil. *Journal of Geotechnical and Geoenvironmental Engineering*, **133**, 257-265.

[5] J. Scherocman, S. Rakowski and K. Uchiyama, Intelligent compaction, does it exist? In *Proceedings of the 52 Annual Conference of the Canadian Technical Asphalt Association*. Polyscience Publications Inc.

[6] H. Thurner and A. Sandström, A new device for instant compaction control. In *Proceedings of International Conference on Compaction*, volume II, 611-4.

[7] D. Adam and F. Kopf , Operational devices for compaction optimization and quality control. *In Geotechnics in Pavement and Railway Design and Construction: Proceedings of the International Seminar on Geotechnics and Railway Design and Construction, Athens, Greece, 16 - 17 December 2004*, 97-106. IOS Press.

[8] T. Yoo and E. Selig, Dynamics of vibratory-roller compaction. *Journal of the Geotechnical Engineering Division*, **105**, 1211-1231.

[9] R. Anderegg and K. Kaufmann, Intelligent compaction with vibratory rollers: Feedback control systems in automatic compaction and compaction control. *Transportation Research Record: Journal of the Transportation Research Board*, **1868**, 124-134.

[10] P.J. van Susante and M.A. Mooney, Capturing nonlinear vibratory roller compactor behavior through lumped parameter modeling. *Journal of Engineering Mechanics*, **134**, 684-693.

[11] R. Rinehart and M. Mooney, Measurement depth of vibratory roller-measured soil stiffness. *Géotechnique*, **59**, 609-619.

[12] D.J. White and M.J. Thompson, Relationships between in situ and roller-integrated compaction measurements for granular soils. *Journal of Geotechnical and Geoenvironmental Engineering*, **134**, 1763-1770.

[13] M.A. Mooney and R.V. Rinehart, In situ soil response to vibratory loading and its relationship to roller-measured soil stiffness. *Journal of Geotechnical and Geoenvironmental Engineering*, **135**, 1022-1031.

[14] A. Puppala, Estimating stiffness of subgrade and unbound materials for pavement design: NCHRP project 20-5.

Paper II

On Moore–Penrose Inverses of quasi-Kronecker Structured Matrices

Daniel K. Heersink & Reinhard Furrer

Paper published in *Linear Algebra and its Applications* (2012), 436, pp. 561–570.

Linear Algebra and its Applications 436 (2012) 561–570

Contents lists available at SciVerse ScienceDirect

Linear Algebra and its Applications

journal homepage: www.elsevier.com/locate/laa

On Moore–Penrose inverses of quasi-Kronecker structured matrices

Daniel K. Heersink [1], Reinhard Furrer [*,2]

Institute of Mathematics, University of Zurich, CH-8057 Zurich, Switzerland

ARTICLE INFO

Article history:
Received 10 May 2011
Accepted 5 July 2011
Available online 4 August 2011

Submitted by R.A. Brualdi

AMS classification:
15A09
15A15
62M30

Keywords:
Block-partitioned matrix
Likelihood
Hierarchical multivariate spatial models

ABSTRACT

The Moore–Penrose inverse and generalized inverse of $\mathbf{A} + \mathbf{X}_1\mathbf{X}_2^*$, where $\mathbf{A}$, $\mathbf{X}_1$, $\mathbf{X}_2$ are complex matrices are given under various assumptions. We use the result to derive the Moore–Penrose inverse and inverse for $\mathrm{bdiag}(\mathbf{A}_k) + \mathbf{u}\mathbf{v}^* \otimes \mathbf{E}$ with p complex matrices $\mathbf{A}_k$, two complex p-vectors $\mathbf{u}$ and $\mathbf{v}$ and a complex matrix $\mathbf{E}$. Such block structured matrices occur in hierarchical modeling of multivariate spatial or space–time Gaussian processes. For the latter we also give expressions of the determinant and of conditional variances.

1. Introduction and notation

The inverse, Moore–Penrose inverse or generalized inverse of $\mathbf{A}+\mathbf{X}_1\mathbf{X}_2^*$, where $\mathbf{A}$ is an $n \times m$ matrix, $\mathbf{X}_1$ is an $n \times r$ matrix, and $\mathbf{X}_2$ is an $m \times r$ matrix, has been discussed extensively in [1–3], and references therein, under various assumptions on the matrices $\mathbf{A}$, $\mathbf{X}_1$ and $\mathbf{X}_2$. For example, if the inverses of $\mathbf{A}$ and $\mathbf{I} + \mathbf{X}_2^*\mathbf{A}^{-1}\mathbf{X}_1$ exist, we have the well-known Sherman–Morrison–Woodbury or Woodbury formula, see [2,3]. If the inverse(s) do not exist, Meyer [4] derives all possible generalized inverses for rank one updates (Sherman–Morrison formula) and Baksalary et al. [5] give alternative representations thereof. Specific cases for higher rank updates are discussed in Riedel [6], Fill and Fishkind [7], Lai and Vemuri [8] and Steerneman and van Perlo-ten Kleij [9]. Further work includes partitioned and block-triangular matrices [10–15].

* Corresponding author.
E-mail addresses: daniel.heersink@math.uzh.ch (D.K. Heersink), reinhard.furrer@math.uzh.ch (R. Furrer).

[1] The contents of this paper are part of the first author's Ph.D. dissertation supervised by the second author.

[2] This work is supported in part by the Swiss National Science Foundation through Project 200021-129782.

0024-3795/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.laa.2011.07.009

In this article we extend the results of Riedel [6] in Section 2 and in Section 3 address the problem of computing "efficiently" the Moore–Penrose inverse of

$$\mathbf{M} = \operatorname{bdiag}(\mathbf{A}_k) + \mathbf{u}\mathbf{v}^* \otimes \mathbf{E}, \tag{1}$$

where $\mathbf{A}_1, \ldots, \mathbf{A}_p$ are arbitrary $n \times m$ complex matrices and $\operatorname{bdiag}(\mathbf{A}_k)$ is the block-diagonal matrix built from $\mathbf{A}_1, \ldots, \mathbf{A}_p$. Without loss of generality, let $n \geqslant m$. $\mathbf{E}$ is an $n \times m$ complex matrix with $\operatorname{rank}(\mathbf{E}) = q$, $0 < q \leqslant m$, $\mathbf{u}$ and $\mathbf{v}$ are complex p-vectors and $\otimes$ is the Kronecker product. Note that (1) is not in classical Kronecker form since different $\mathbf{A}_k$ are involved. We term this "quasi-Kronecker" structure.

The key observation is that (1) can be written as

$$\mathbf{M} = \operatorname{bdiag}(\mathbf{A}_k) + (\mathbf{u} \otimes \mathbf{E}_\mathbf{u})(\mathbf{v} \otimes \mathbf{E}_\mathbf{v})^* \tag{2}$$

for any $n \times r$ and $m \times r$ matrices $\mathbf{E}_\mathbf{u}$ and $\mathbf{E}_\mathbf{v}$ such that $\mathbf{E}_\mathbf{u}\mathbf{E}_\mathbf{v}^* = \mathbf{E}$. As an example, assume that the inverses of $\mathbf{A}_k$, $k = 1, \ldots, p$, and $\mathbf{I}_r + (\mathbf{v} \otimes \mathbf{E}_\mathbf{v})^* \operatorname{bdiag}(\mathbf{A}_k^{-1})(\mathbf{u} \otimes \mathbf{E}_\mathbf{u})$ exist, $\mathbf{I}_r$ being the identity matrix of dimension r. Then the inverse of $\mathbf{M}$ is [2]

$$\begin{aligned}\mathbf{M}^{-1} = {} & \operatorname{bdiag}(\mathbf{A}_k^{-1}) - \operatorname{bdiag}(\mathbf{A}_k^{-1})(\mathbf{u} \otimes \mathbf{E}_\mathbf{u}) \\ & \times \left(\mathbf{I}_r + (\mathbf{v} \otimes \mathbf{E}_\mathbf{v})^* \operatorname{bdiag}(\mathbf{A}_k^{-1})(\mathbf{u} \otimes \mathbf{E}_\mathbf{u})\right)^{-1} (\mathbf{v} \otimes \mathbf{E}_\mathbf{v})^* \operatorname{bdiag}(\mathbf{A}_k^{-1});\end{aligned} \tag{3}$$

hence based on the calculation of p $n \times n$ inverses and one $r \times r$ inverse, instead of one $np \times np$ inverse.

We use the following notation. Matrices and vectors are bold face; $\mathbf{0}$ is the zero matrix, where the dimensions are given by the context; $\bar{c}$ is the complex conjugate of c; $\mathbf{A}^T$ and $\mathbf{A}^*$ denote the transpose and conjugate transpose of $\mathbf{A}$; $\mathbf{A}^+$ and $\mathbf{A}^-$ are the Moore–Penrose inverse and generalized inverse; $\operatorname{block}(\mathbf{N}_{ij})$ denotes a partitioned matrix composed of the blocks $\mathbf{N}_{ij}$; $\operatorname{bcol}(\mathbf{N}_i)$ and $\operatorname{brow}(\mathbf{N}_j)$ denote a column and row partitioned matrix composed of the blocks $\mathbf{N}_i$ and $\mathbf{N}_j$, respectively. The column space of $\mathbf{A}$ is denoted by $R(\mathbf{A})$. The orthogonal complement to $R(\mathbf{A})$ is denoted by $R(\mathbf{A})^\perp$. The unique, orthogonal projector onto $R(\mathbf{A})$ is $\mathbf{P}_\mathbf{A} = \mathbf{A}\mathbf{A}^+$ and onto $R(\mathbf{A})^\perp$ is $\mathbf{Q}_\mathbf{A} = \mathbf{I}_n - \mathbf{A}\mathbf{A}^+$. The unique, orthogonal projector onto $R(\mathbf{A}^*)$ is $\mathbf{P}_{\mathbf{A}^*} = \mathbf{A}^+\mathbf{A}$ and onto $R(\mathbf{A}^*)^\perp$ is $\mathbf{Q}_{\mathbf{A}^*} = \mathbf{I}_m - \mathbf{A}^+\mathbf{A}$.

2. Main results

We first consider the Moore–Penrose inverse of $\mathbf{A} + \mathbf{B}$, where $\mathbf{A}$ and $\mathbf{B}$ are arbitrary $n \times m$ complex matrices. Without loss of generality, let $n \geqslant m$. Assume that $\mathbf{B}$ is of rank $0 < q \leqslant m$ and let the singular value decomposition (SVD) of $\mathbf{B} = \tilde{\mathbf{U}}\tilde{\mathbf{D}}\tilde{\mathbf{V}}^*$. Without loss of generality, we assume that the diagonal entries of $\tilde{\mathbf{D}}$ are sorted in decreasing order. Let $\tilde{\mathbf{U}}_{1:q}$ and $\tilde{\mathbf{V}}_{1:q}$ be the first q columns of $\tilde{\mathbf{U}}$ and $\tilde{\mathbf{V}}$ and let $\tilde{\mathbf{D}}_{1:q}^{1/2}$ be the diagonal square matrix with entries $\tilde{\mathbf{D}}_{ii}^{1/2}$, $i = 1, \ldots, q$. Finally, let $\mathbf{U} = \tilde{\mathbf{U}}_{1:q}\tilde{\mathbf{D}}_{1:q}^{1/2}$ and $\mathbf{V} = \tilde{\mathbf{V}}_{1:q}\tilde{\mathbf{D}}_{1:q}^{1/2}$, i.e., $\mathbf{B} = \mathbf{U}\mathbf{V}^*$. We now decompose $\mathbf{U}$ and $\mathbf{V}$ into $\mathbf{W}_1 + \mathbf{Z}_1$ and $\mathbf{W}_2 + \mathbf{Z}_2$, where the columns of $\mathbf{W}_1$ are contained in the column space of $\mathbf{A}$, $R(\mathbf{A})$, and the columns of $\mathbf{Z}_1$ are orthogonal to $R(\mathbf{A})$. Similarly, the columns of $\mathbf{W}_2$ and $\mathbf{Z}_2$ are in $R(\mathbf{A}^*)$ and $R(\mathbf{A}^*)^\perp$, respectively. The first theorem gives the Moore–Penrose inverse of $\mathbf{M} = \mathbf{A} + (\mathbf{W}_1 + \mathbf{Z}_1)(\mathbf{W}_2 + \mathbf{Z}_2)^*$ under various assumptions on $\mathbf{A}$, $\mathbf{W}_\ell$ and $\mathbf{Z}_\ell$, $\ell = 1, 2$.

Theorem 1. *Let $\mathbf{A}$, $\mathbf{B} = \mathbf{U}\mathbf{V}^*$ be $n \times m$ complex matrices. Let $\mathbf{W}_1 = \mathbf{P}_\mathbf{A}\mathbf{U}$, $\mathbf{Z}_1 = \mathbf{Q}_\mathbf{A}\mathbf{U}$, $\mathbf{W}_2 = \mathbf{P}_{\mathbf{A}^*}\mathbf{V}$ and $\mathbf{Z}_2 = \mathbf{Q}_{\mathbf{A}^*}\mathbf{V}$. Define $\mathbf{Y} = \mathbf{I}_q + \mathbf{W}_2^*\mathbf{A}^+\mathbf{W}_1$, $\mathbf{K}_1 = \mathbf{A}^+\mathbf{W}_1$, and $\mathbf{K}_2 = \mathbf{W}_2^*\mathbf{A}^+$. Under the assumptions detailed below, the Moore–Penrose inverse of $\mathbf{M} = \mathbf{A} + \mathbf{U}\mathbf{V}^*$ is*

$$\mathbf{M}^+ = (\mathbf{A} + \mathbf{U}\mathbf{V}^*)^+ = \mathbf{A}^+ - \mathbf{N},$$

where

A. $\mathbf{N} = \mathbf{Z}_2^{+*}\mathbf{K}_2 + \mathbf{K}_1\mathbf{Z}_1^+ - \mathbf{Z}_2^{+*}\mathbf{Y}\mathbf{Z}_1^+$, *if* $\operatorname{rank}(\mathbf{Z}_1) = \operatorname{rank}(\mathbf{Z}_2) = q$.
B. $\mathbf{N} = \mathbf{K}_1\mathbf{Z}_1^+ + \mathbf{A}^+\mathbf{K}_2^+\mathbf{K}_2$, *if* $\operatorname{rank}(\mathbf{Z}_1) = q$, $\mathbf{Z}_2 = \mathbf{0}$ *and* $\mathbf{Y} = \mathbf{0}$.

C. $\mathbf{N} = \mathbf{Z}_2^{+*}\mathbf{K}_2 + \mathbf{K}_1\mathbf{K}_1^{+}\mathbf{A}^{+}$, *if* $\mathbf{Z}_1 = \mathbf{0}$, $\operatorname{rank}(\mathbf{Z}_2) = q$ *and* $\mathbf{Y} = \mathbf{0}$.
D. $\mathbf{N} = \mathbf{K}_1\mathbf{K}_1^{+}\mathbf{A}^{+} + \mathbf{A}^{+}\mathbf{K}_2^{+}\mathbf{K}_2 - \mathbf{K}_1\mathbf{K}_1^{+}\mathbf{A}^{+}\mathbf{K}_2^{+}\mathbf{K}_2$, *if* $\mathbf{Z}_1 = \mathbf{0}$, $\mathbf{Z}_2 = \mathbf{0}$ *and* $\mathbf{Y} = \mathbf{0}$.
E. $\mathbf{N} = \mathbf{K}_1\mathbf{Y}^{-1}\mathbf{K}_2$, *if* $\mathbf{Z}_1 = \mathbf{0}$, $\mathbf{Z}_2 = \mathbf{0}$ *and* $\operatorname{rank}(\mathbf{Y}) = q$.
F. $\mathbf{N} = (\mathbf{S}_1\mathbf{Z}_1 + \mathbf{K}_1)(\mathbf{Y} + \mathbf{W}_2^{*}\mathbf{S}_1\mathbf{Z}_1)^{-1}\mathbf{T}_1 - \mathbf{S}_1$, *with* $\mathbf{S}_1 = \mathbf{A}^{+}\mathbf{K}_2^{*}\mathbf{Y}^{*-1}\mathbf{Z}_1^{*}$ *and* $\mathbf{T}_1 = \mathbf{W}_2^{*}\mathbf{S}_1 + \mathbf{K}_2$, *if* $\mathbf{Z}_2 = \mathbf{0}$ *and* $\operatorname{rank}(\mathbf{Y}) = q$.
G. $\mathbf{N} = \mathbf{T}_2(\mathbf{Y} + \mathbf{Z}_2^{*}\mathbf{S}_2\mathbf{W}_1)^{-1}(\mathbf{Z}_2^{*}\mathbf{S}_2 + \mathbf{K}_2) - \mathbf{S}_2$, *with* $\mathbf{S}_2 = \mathbf{Z}_2\mathbf{Y}^{*-1}\mathbf{K}_1^{*}\mathbf{A}^{+}$ *and* $\mathbf{T}_2 = \mathbf{S}_2\mathbf{W}_1 + \mathbf{K}_1$, *if* $\mathbf{Z}_1 = \mathbf{0}$ *and* $\operatorname{rank}(\mathbf{Y}) = q$.

The theorem can be shown by verifying the four Moore–Penrose conditions [16] (I) $\mathbf{M}\mathbf{M}^{+}\mathbf{M} = \mathbf{M}$, (II) $\mathbf{M}^{+}\mathbf{M}\mathbf{M}^{+} = \mathbf{M}^{+}$, (III) $\mathbf{M}\mathbf{M}^{+}$ and (IV) $\mathbf{M}^{+}\mathbf{M}$ are Hermitian. We further need the following useful fact.

Lemma 2. *With the notation of Theorem 1.F and 1.G,* $(\mathbf{Y}+\mathbf{W}_2^{*}\mathbf{S}_1\mathbf{Z}_1)^{-1}$ *and* $(\mathbf{Y}+\mathbf{Z}_2^{*}\mathbf{S}_2\mathbf{W}_1)^{-1}$, *respectively, exist.*

Proof. Define $|\mathbf{A}|$ to be the determinant of $\mathbf{A}$. $(\mathbf{Y} + \mathbf{W}_2^{*}\mathbf{S}_1\mathbf{Z}_1)^{-1}$ exists $\iff |\mathbf{Y} + \mathbf{W}_2^{*}\mathbf{S}_1\mathbf{Z}_1| \neq 0$. $|\mathbf{Y} + \mathbf{W}_2^{*}\mathbf{S}_1\mathbf{Z}_1| = |\mathbf{Y} + \mathbf{K}_2\mathbf{K}_2^{*}\mathbf{Y}^{*-1}\mathbf{Z}_1^{*}\mathbf{Z}_1| = |\mathbf{Y}||\mathbf{I}_n + \mathbf{K}_2^{*}\mathbf{Y}^{*-1}\mathbf{Z}_1^{*}\mathbf{Z}_1\mathbf{Y}^{-1}\mathbf{K}_2|$ [13, p. 416]. Let $\operatorname{rank}(\mathbf{Z}_1) = q'$, $0 < q' \leqslant q$. Thus $\mathbf{Z}_1^{*}\mathbf{Z}_1$ is positive semi-definite. Therefore, $\mathbf{K}_2^{*}\mathbf{Y}^{*-1}\mathbf{Z}_1^{*}\mathbf{Z}_1\mathbf{Y}^{-1}\mathbf{K}_2$ is also positive semi-definite [13, p. 213]. Which implies $\mathbf{I}_n + \mathbf{K}_2^{*}\mathbf{Y}^{*-1}\mathbf{Z}_1^{*}\mathbf{Z}_1\mathbf{Y}^{-1}\mathbf{K}_2$ is positive definite. Therefore, $|\mathbf{I}_n + \mathbf{K}_2^{*}\mathbf{Y}^{*-1}\mathbf{Z}_1^{*}\mathbf{Z}_1\mathbf{Y}^{-1}\mathbf{K}_2| > 0$ and $(\mathbf{Y} + \mathbf{W}_2^{*}\mathbf{S}_1\mathbf{Z}_1)^{-1}$ exists. Proof of the existence of $(\mathbf{Y} + \mathbf{Z}_2^{*}\mathbf{S}_2\mathbf{W}_1)^{-1}$ is similar. □

Proof of Theorem 1. All four Moore–Penrose conditions are verified for cases A and F. $\mathbf{M}\mathbf{M}^{+}$ and $\mathbf{M}^{+}\mathbf{M}$ are provided for the remaining cases and the calculations are similar to cases A and F. All cases can be shown by direct verification and making repeated use of several identities listed in the proofs of A and F.

A. We have $\mathbf{M} = \mathbf{A} + (\mathbf{W}_1 + \mathbf{Z}_1)(\mathbf{W}_2 + \mathbf{Z}_2)^{*}$ and

$$\mathbf{M}\mathbf{M}^{+} = (\mathbf{A} + (\mathbf{W}_1 + \mathbf{Z}_1)(\mathbf{W}_2 + \mathbf{Z}_2)^{*})(\mathbf{A}^{+} - \mathbf{Z}_2^{+*}\mathbf{K}_2 - \mathbf{K}_1\mathbf{Z}_1^{+} + \mathbf{Z}_2^{+*}\mathbf{Y}\mathbf{Z}_1^{+}).$$

Expanding the equality and using the identities (a) $\mathbf{A}\mathbf{Z}_2^{+*} = \mathbf{0}$, (b) $\mathbf{Z}_2^{*}\mathbf{A}^{+} = \mathbf{0}$, (c) $\mathbf{W}_2^{*}\mathbf{Z}_2^{+*} = \mathbf{0}$, and (d) $\mathbf{A}\mathbf{K}_1 = \mathbf{W}_1$, yields

$$\begin{aligned}\mathbf{M}\mathbf{M}^{+} = {} & \mathbf{A}\mathbf{A}^{+} - \mathbf{W}_1\mathbf{Z}_1^{+} + (\mathbf{W}_1 + \mathbf{Z}_1)\mathbf{K}_2 - (\mathbf{W}_1 + \mathbf{Z}_1)\mathbf{Z}_2^{*}\mathbf{Z}_2^{+*}\mathbf{K}_2 \\ & - (\mathbf{W}_1 + \mathbf{Z}_1)\mathbf{W}_2^{*}\mathbf{K}_1\mathbf{Z}_1^{+} + (\mathbf{W}_1 + \mathbf{Z}_1)\mathbf{Z}_2^{*}\mathbf{Z}_2^{+*}\mathbf{Y}\mathbf{Z}_1^{+}.\end{aligned}$$

Further, using the identities (e) $\mathbf{Z}_2^{*}\mathbf{Z}_2^{+*} = \mathbf{I}_q$ and (f) $\mathbf{W}_2^{*}\mathbf{K}_1 = \mathbf{Y} - \mathbf{I}_q$, and simplifying the expression we get

$$\mathbf{M}\mathbf{M}^{+} = \mathbf{A}\mathbf{A}^{+} + \mathbf{Z}_1\mathbf{Z}_1^{+}.$$

We can expand $\mathbf{M}^{+}\mathbf{M}$ in a similar way.

$$\mathbf{M}^{+}\mathbf{M} = (\mathbf{A}^{+} - \mathbf{Z}_2^{+*}\mathbf{K}_2 - \mathbf{K}_1\mathbf{Z}_1^{+} + \mathbf{Z}_2^{+*}\mathbf{Y}\mathbf{Z}_1^{+})(\mathbf{A} + (\mathbf{W}_1 + \mathbf{Z}_1)(\mathbf{W}_2 + \mathbf{Z}_2)^{*}).$$

Expanding the equality and using the identities (g) $\mathbf{A}^{+}\mathbf{Z}_1 = \mathbf{0}$, (h) $\mathbf{Z}_1^{+}\mathbf{A} = \mathbf{0}$, (i) $\mathbf{Z}_1^{+}\mathbf{W}_1 = \mathbf{0}$, and (j) $\mathbf{K}_2\mathbf{A} = \mathbf{W}_2^{*}$, yields

$$\begin{aligned}\mathbf{M}^{+}\mathbf{M} = {} & \mathbf{A}^{+}\mathbf{A} + \mathbf{K}_1(\mathbf{W}_2 + \mathbf{Z}_2)^{*} - \mathbf{Z}_2^{+*}\mathbf{W}_2^{*} - \mathbf{Z}_2^{+*}\mathbf{K}_2\mathbf{W}_1(\mathbf{W}_2 + \mathbf{Z}_2)^{*} \\ & - \mathbf{K}_1\mathbf{Z}_1^{+}\mathbf{Z}_1(\mathbf{W}_2 + \mathbf{Z}_2)^{*} + \mathbf{Z}_2^{+*}\mathbf{Y}\mathbf{Z}_1^{+}\mathbf{Z}_1(\mathbf{W}_2 + \mathbf{Z}_2)^{*}.\end{aligned}$$

Further, using the identities (k) $\mathbf{Z}_1^{+}\mathbf{Z}_1 = \mathbf{I}_q$ and (l) $\mathbf{K}_2\mathbf{W}_1 = \mathbf{Y} - \mathbf{I}_q$, and simplifying the expression we get

$$\mathbf{M}^+\mathbf{M} = \mathbf{A}^+\mathbf{A} + \mathbf{Z}_2^{+*}\mathbf{Z}_2^*.$$

Therefore, $\mathbf{M}\mathbf{M}^+$ and $\mathbf{M}^+\mathbf{M}$ are Hermitian. Thus, Moore–Penrose conditions (III) and (IV) are satisfied. Note that identities (e) and (k) are only valid under the full rank assumption of $\mathbf{Z}_2$ and $\mathbf{Z}_1$, respectively. Now, using identities (a), (b), and (c) $\mathbf{M}\mathbf{M}^+\mathbf{M}$ expands and simplifies to

$$\begin{aligned}\mathbf{M}\mathbf{M}^+\mathbf{M} &= (\mathbf{A} + (\mathbf{W}_1 + \mathbf{Z}_1)(\mathbf{W}_2 + \mathbf{Z}_2)^*)(\mathbf{A}^+\mathbf{A} + \mathbf{Z}_2^{+*}\mathbf{Z}_2^*)\\ &= \mathbf{A}\mathbf{A}^+\mathbf{A} + (\mathbf{W}_1 + \mathbf{Z}_1)\mathbf{W}_2^*\mathbf{A}^+\mathbf{A} + (\mathbf{W}_1 + \mathbf{Z}_1)\mathbf{Z}_2^*\mathbf{Z}_2^{+*}\mathbf{Z}_2^*.\end{aligned}$$

Using (I) for $\mathbf{A}$ and $\mathbf{Z}_2$ and noting that $\mathbf{W}_2^*\mathbf{A}^+\mathbf{A} = \mathbf{W}_2^*$, this simplifies to

$$\mathbf{M}\mathbf{M}^+\mathbf{M} = \mathbf{A} + (\mathbf{W}_1 + \mathbf{Z}_1)\mathbf{W}_2^* + (\mathbf{W}_1 + \mathbf{Z}_1)\mathbf{Z}_2^* = \mathbf{M}.$$

Thus Moore–Penrose condition (I) is satisfied. Utilizing identities (g), (h), and (i), $\mathbf{M}^+\mathbf{M}\mathbf{M}^+$ expands and simplifies to

$$\begin{aligned}\mathbf{M}^+\mathbf{M}\mathbf{M}^+ &= (\mathbf{A}^+ - \mathbf{Z}_2^{+*}\mathbf{K}_2 - \mathbf{K}_1\mathbf{Z}_1^+ + \mathbf{Z}_2^{+*}\mathbf{Y}\mathbf{Z}_1^+)(\mathbf{A}\mathbf{A}^+ + \mathbf{Z}_1\mathbf{Z}_1^+)\\ &= \mathbf{A}^+\mathbf{A}\mathbf{A}^+ - \mathbf{Z}_2^{+*}\mathbf{K}_2\mathbf{A}\mathbf{A}^+ - \mathbf{K}_1\mathbf{Z}_1^+\mathbf{Z}_1\mathbf{Z}_1^+ + \mathbf{Z}_2^{+*}\mathbf{Y}\mathbf{Z}_1^+\mathbf{Z}_1\mathbf{Z}_1^+.\end{aligned}$$

Using (II) for $\mathbf{A}$ and $\mathbf{Z}_1$ and noting that $\mathbf{K}_2\mathbf{A}\mathbf{A}^+ = \mathbf{K}_2$, this simplifies to

$$\mathbf{M}^+\mathbf{M}\mathbf{M}^+ = \mathbf{A}^+ - \mathbf{Z}_2^{+*}\mathbf{K}_2 - \mathbf{K}_1\mathbf{Z}_1^+ + \mathbf{Z}_2^{+*}\mathbf{Y}\mathbf{Z}_1^+ = \mathbf{M}^+.$$

Thus Moore–Penrose condition (II) is satisfied and $\mathbf{M}^+$ is the Moore–Penrose inverse of $\mathbf{M}$.

B. We have $\mathbf{M} = \mathbf{A} + (\mathbf{W}_1 + \mathbf{Z}_1)\mathbf{W}_2^*$ and

$$\mathbf{M}\mathbf{M}^+ = \mathbf{A}\mathbf{A}^+ + \mathbf{Z}_1\mathbf{Z}_1^+ - \mathbf{K}_2^+\mathbf{K}_2 \quad \text{and} \quad \mathbf{M}^+\mathbf{M} = \mathbf{A}^+\mathbf{A}.$$

C. We have $\mathbf{M} = \mathbf{A} + \mathbf{W}_1(\mathbf{W}_2 + \mathbf{Z}_2)^*$ and

$$\mathbf{M}\mathbf{M}^+ = \mathbf{A}\mathbf{A}^+ \quad \text{and} \quad \mathbf{M}^+\mathbf{M} = \mathbf{A}^+\mathbf{A} + \mathbf{Z}_2^{+*}\mathbf{Z}_2^* - \mathbf{K}_1\mathbf{K}_1^+.$$

D. We have $\mathbf{M} = \mathbf{A} + \mathbf{W}_1\mathbf{W}_2^*$ and

$$\mathbf{M}\mathbf{M}^+ = \mathbf{A}\mathbf{A}^+ - \mathbf{K}_2^+\mathbf{K}_2 \quad \text{and} \quad \mathbf{M}^+\mathbf{M} = \mathbf{A}^+\mathbf{A} - \mathbf{K}_1\mathbf{K}_1^+.$$

E. The inverse of $\mathbf{Y}$ exists by assumption, $\mathbf{M} = \mathbf{A} + \mathbf{W}_1\mathbf{W}_2^*$ and

$$\begin{aligned}\mathbf{M}\mathbf{M}^+ &= \mathbf{A}\mathbf{A}^+ + \mathbf{W}_1\mathbf{K}_2 - \mathbf{W}_1(\mathbf{I}_q + \mathbf{W}_2^*\mathbf{A}^+\mathbf{W}_1)\mathbf{Y}^{-1}\mathbf{K}_2 = \mathbf{A}\mathbf{A}^+,\\ \mathbf{M}^+\mathbf{M} &= \mathbf{A}^+\mathbf{A} + \mathbf{K}_1\mathbf{W}_2^* - \mathbf{K}_1\mathbf{Y}^{-1}(\mathbf{I}_q + \mathbf{W}_2^*\mathbf{A}^+\mathbf{W}_1)\mathbf{W}_2^* = \mathbf{A}^+\mathbf{A}.\end{aligned}$$

F. Note that the inverses of $\mathbf{Y} + \mathbf{W}_2^*\mathbf{S}_1\mathbf{Z}_1$ and $\mathbf{Y}$ exist by assumption (Lemma 2). For ease of notation, let $\mathbf{R}_1 = \mathbf{Y} + \mathbf{W}_2^*\mathbf{S}_1\mathbf{Z}_1$ and $\widetilde{\mathbf{S}}_1 = \mathbf{S}_1\mathbf{Z}_1 + \mathbf{K}_1$. Note that $\mathbf{M} = \mathbf{A} + (\mathbf{W}_1 + \mathbf{Z}_1)\mathbf{W}_2^*$ and

$$\mathbf{M}^+\mathbf{M} = (\mathbf{A}^+ - \widetilde{\mathbf{S}}_1\mathbf{R}_1^{-1}\mathbf{T}_1 + \mathbf{S}_1)(\mathbf{A} + (\mathbf{W}_1 + \mathbf{Z}_1)\mathbf{W}_2^*).$$

Expanding the equality and using the identities (g), (m) $\mathbf{T}_1\mathbf{A} = \mathbf{W}_2^*$, (n) $\mathbf{T}_1(\mathbf{W}_1 + \mathbf{Z}_1) = \mathbf{R}_1 - \mathbf{I}_q$, (o) $\mathbf{Z}_1^*\mathbf{A} = \mathbf{0}$, and (p) $\mathbf{Z}_1^*\mathbf{W}_1 = \mathbf{0}$, yields

$$\begin{aligned}\mathbf{M}^+\mathbf{M} &= \mathbf{A}^+\mathbf{A} + \mathbf{K}_1\mathbf{W}_2^* - \widetilde{\mathbf{S}}_1\mathbf{R}_1^{-1}\mathbf{W}_2^* - \widetilde{\mathbf{S}}_1\mathbf{R}_1^{-1}(\mathbf{R}_1 - \mathbf{I}_q)\mathbf{W}_2^* + \mathbf{S}_1\mathbf{Z}_1\mathbf{W}_2^*\\ &= \mathbf{A}^+\mathbf{A} + \mathbf{K}_1\mathbf{W}_2^* - \widetilde{\mathbf{S}}_1\mathbf{W}_2^* + \mathbf{S}_1\mathbf{Z}_1\mathbf{W}_2^*\\ &= \mathbf{A}^+\mathbf{A} + \mathbf{K}_1\mathbf{W}_2^* - (\mathbf{S}_1\mathbf{Z}_1 + \mathbf{K}_1)\mathbf{W}_2^* + \mathbf{S}_1\mathbf{Z}_1\mathbf{W}_2^*\\ &= \mathbf{A}^+\mathbf{A} + \mathbf{K}_1\mathbf{W}_2^* - \mathbf{S}_1\mathbf{Z}_1\mathbf{W}_2^* - \mathbf{K}_1\mathbf{W}_2^* + \mathbf{S}_1\mathbf{Z}_1\mathbf{W}_2^* = \mathbf{A}^+\mathbf{A}.\end{aligned}$$

Thus, (IV) is satisfied. Using identity (q) $\mathbf{W}_2^*\widetilde{\mathbf{S}}_1 = \mathbf{R}_1 - \mathbf{I}_q$, after expansion we get

$$\begin{aligned}\mathbf{M}\mathbf{M}^+ &= (\mathbf{A} + (\mathbf{W}_1 + \mathbf{Z}_1)\mathbf{W}_2^*)(\mathbf{A}^+ - \widetilde{\mathbf{S}}_1\mathbf{R}_1^{-1}\mathbf{T}_1 + \mathbf{S}_1)\\ &= \mathbf{A}\mathbf{A}^+ - \mathbf{A}\widetilde{\mathbf{S}}_1\mathbf{R}_1^{-1}\mathbf{T}_1 + \mathbf{A}\mathbf{S}_1 + (\mathbf{W}_1 + \mathbf{Z}_1)\mathbf{K}_2\\ &\quad - (\mathbf{W}_1 + \mathbf{Z}_1)(\mathbf{R}_1 - \mathbf{I}_q)\mathbf{R}_1^{-1}\mathbf{T}_1 + (\mathbf{W}_1 + \mathbf{Z}_1)\mathbf{W}_2^*\mathbf{S}_1\\ &= \mathbf{A}\mathbf{A}^+ - \mathbf{A}\widetilde{\mathbf{S}}_1\mathbf{R}_1^{-1}\mathbf{T}_1 + \mathbf{A}\mathbf{S}_1 + (\mathbf{W}_1 + \mathbf{Z}_1)\mathbf{R}_1^{-1}\mathbf{T}_1\\ &= \mathbf{A}\mathbf{A}^+ + \mathbf{A}\mathbf{S}_1 - (\mathbf{A}\widetilde{\mathbf{S}}_1 - \mathbf{W}_1 - \mathbf{Z}_1)\mathbf{R}_1^{-1}\mathbf{T}_1.\end{aligned}$$

Note that $\operatorname{rank}(\mathbf{Z}_2) = 0 \Rightarrow \operatorname{rank}(\mathbf{K}_2) = q$, thus $(\mathbf{K}_2\mathbf{K}_2^*)^{-1}$ exists. Using identities (r) $\mathbf{Z}_1 = (\mathbf{T}_1 - \mathbf{K}_2)^*(\mathbf{K}_2\mathbf{K}_2^*)^{-1}\mathbf{Y}$, (s) $\mathbf{A}\widetilde{\mathbf{S}}_1 - \mathbf{W}_1 = \mathbf{K}_2^*\mathbf{Y}^{*-1}\mathbf{Z}_1^*\mathbf{Z}_1$, and (t) $\mathbf{T}_1\mathbf{T}_1^* = \mathbf{R}_1\mathbf{Y}^{-1}\mathbf{K}_2\mathbf{K}_2^*$, this can be rewritten as

$$\begin{aligned}\mathbf{M}\mathbf{M}^+ &= \mathbf{A}\mathbf{A}^+ + \mathbf{K}_2^*\mathbf{Y}^{*-1}\mathbf{Z}_1^* + \mathbf{T}_1^*(\mathbf{T}_1\mathbf{T}_1^*)^{-1}\mathbf{T}_1 - \mathbf{K}_2^*(\mathbf{K}_2\mathbf{K}_2^*)^{-1}\mathbf{T}_1\\ &= \mathbf{A}\mathbf{A}^+ + \mathbf{T}_1^*\mathbf{T}_1^{+*} + \mathbf{K}_2^*(\mathbf{Y}^{*-1}\mathbf{Z}_1^* - \mathbf{Y}^{*-1}\mathbf{Z}_1^* - (\mathbf{K}_2\mathbf{K}_2^*)^{-1}\mathbf{K}_2)\\ &= \mathbf{A}\mathbf{A}^+ + \mathbf{T}_1^*\mathbf{T}_1^{+*} - \mathbf{K}_2^+\mathbf{K}_2.\end{aligned}$$

Thus, (III) is satisfied. Noting again that $\mathbf{W}_2^*\mathbf{A}^+\mathbf{A} = \mathbf{W}_2^*$ and using (I) for $\mathbf{A}$,

$$\begin{aligned}\mathbf{M}\mathbf{M}^+\mathbf{M} &= (\mathbf{A} + (\mathbf{W}_1 + \mathbf{Z}_1)\mathbf{W}_2^*)\mathbf{A}^+\mathbf{A}\\ &= \mathbf{A}\mathbf{A}^+\mathbf{A} + (\mathbf{W}_1 + \mathbf{Z}_1)\mathbf{W}_2^*\mathbf{A}^+\mathbf{A} = \mathbf{M}.\end{aligned}$$

Thus, (I) is satisfied. Also noting that $\mathbf{A}^+\mathbf{A}\widetilde{\mathbf{S}}_1 = \widetilde{\mathbf{S}}_1$ and utilizing (II) for $\mathbf{A}$ we show that Moore–Penrose condition (II) is satisfied:

$$\begin{aligned}\mathbf{M}^+\mathbf{M}\mathbf{M}^+ &= \mathbf{A}^+\mathbf{A}(\mathbf{A}^+ - \widetilde{\mathbf{S}}_1\mathbf{R}_1^{-1}\mathbf{T}_1 + \mathbf{A}^+\mathbf{K}_2^*\mathbf{Y}^{*-1}\mathbf{Z}_1^*)\\ &= \mathbf{A}^+\mathbf{A}\mathbf{A}^+ - \mathbf{A}^+\mathbf{A}\widetilde{\mathbf{S}}_1\mathbf{R}_1^{-1}\mathbf{T}_1 + \mathbf{A}^+\mathbf{A}\mathbf{A}^+\mathbf{K}_2^*\mathbf{Y}^{*-1}\mathbf{Z}_1^* = \mathbf{M}^+.\end{aligned}$$

Thus $\mathbf{M}^+$ is the Moore–Penrose inverse of $\mathbf{M}$.

G. The case $\mathbf{M} = \mathbf{A} + \mathbf{W}_1(\mathbf{W}_2 + \mathbf{Z}_2)^*$ is shown similarly to F and

$$\mathbf{M}\mathbf{M}^+ = \mathbf{A}\mathbf{A}^+ \quad \text{and} \quad \mathbf{M}^+\mathbf{M} = \mathbf{A}^+\mathbf{A} - \mathbf{K}_1\mathbf{K}_1^+ + \mathbf{T}_2\mathbf{T}_2^+. \qquad \square$$

Remark 1

(1) Theorem 1 states explicitly the Moore–Penrose inverses for all full or zero rank cases of $\mathbf{Z}_1$, $\mathbf{Z}_2$, and $\mathbf{Y}$, along the lines of Meyer's [4] list of rank one updates. Riedel's [6] result is a particular case of Theorem 1.A.

(2) The full rank conditions on $\mathbf{Z}_1, \mathbf{Z}_2$ and $\mathbf{Y}$ can be relaxed as follows. For part A we require $\mathbf{Q}_{\mathbf{Z}_2^*}\mathbf{W}_2^* = \mathbf{0}$, $\mathbf{Q}_{\mathbf{Z}_2^*}\mathbf{Z}_1^+ = \mathbf{0}$, $\mathbf{W}_1\mathbf{Q}_{\mathbf{Z}_1^*} = \mathbf{0}$, and $\mathbf{Z}_2^{+*}\mathbf{Q}_{\mathbf{Z}_1^*} = \mathbf{0}$, since without the full rank assumptions on $\mathbf{Z}_1$ and $\mathbf{Z}_2$ identities (e) and (k) do not necessarily hold. For parts B and C, it is sufficient to assume $\mathbf{W}_1\mathbf{Q}_{\mathbf{Z}_1^*} = \mathbf{0}$ and $\mathbf{Q}_{\mathbf{Z}_2^*}\mathbf{W}_2^* = \mathbf{0}$, respectively, since identities (k) and (e), respectively, do not necessarily hold. Finally for part E, the condition $\operatorname{rank}(\mathbf{Y}) = q$ can be relaxed towards the conditions that $\mathbf{W}_1\mathbf{Q}_{\mathbf{Y}}\mathbf{K}_2 = \mathbf{0}$ and $\mathbf{K}_1\mathbf{Q}_{\mathbf{Y}^*}\mathbf{W}_2^* = \mathbf{0}$. All of these conditions impose constraints on the row spaces of the matrices $\mathbf{Z}_\ell$ and $\mathbf{W}_\ell$, $\ell = 1, 2$.

(3) In the case of symmetric matrices, parts B and C are irrelevant since that implies $\operatorname{rank}(\mathbf{Z}_1) = \operatorname{rank}(\mathbf{Z}_2)$.

(4) The construction of $\mathbf{U}$ and $\mathbf{V}$ via a SVD is not required. Any decomposition based on two rank q matrices is sufficient.

If only an arbitrary generalized inverse is needed (i.e., only satisfying the first Penrose condition $\mathbf{M}\mathbf{M}^{-}\mathbf{M} = \mathbf{M}$), we can weaken the assumptions of Theorem 1.A–D and simplify the expressions of the Moore–Penrose inverses for cases B to D.

Theorem 3. *Let* $\mathbf{A}$, $\mathbf{B} = \mathbf{U}\mathbf{V}^*$ *be* $n \times m$ *complex matrices. Let* $\mathbf{W}_1 = \mathbf{P}_{\mathbf{A}}\mathbf{U}$, $\mathbf{Z}_1 = \mathbf{Q}_{\mathbf{A}}\mathbf{U}$, $\mathbf{W}_2 = \mathbf{P}_{\mathbf{A}^*}\mathbf{V}$ *and* $\mathbf{Z}_2 = \mathbf{Q}_{\mathbf{A}^*}\mathbf{V}$. *Define* $\mathbf{Y} = \mathbf{I}_q + \mathbf{W}_2^*\mathbf{A}^+\mathbf{W}_1$, $\mathbf{K}_1 = \mathbf{A}^+\mathbf{W}_1$, *and* $\mathbf{K}_2 = \mathbf{W}_2^*\mathbf{A}^+$. *Under the assumptions detailed below, a generalized inverse of* $\mathbf{M} = \mathbf{A} + \mathbf{U}\mathbf{V}^*$ *is*

$$\mathbf{M}^{-} = (\mathbf{A} + \mathbf{U}\mathbf{V}^*)^{-} = \mathbf{A}^+ - \mathbf{N},$$

where

A. $\mathbf{N} = \mathbf{Z}_2^{+*}\mathbf{K}_2 + \mathbf{K}_1\mathbf{Z}_1^+ - \mathbf{Z}_2^{+*}\mathbf{Y}\mathbf{Z}_1^+$, *if* $\operatorname{rank}(\mathbf{Z}_1) = q$ *and* $0 \leqslant \operatorname{rank}(\mathbf{Z}_2) \leqslant q$ *or* $0 \leqslant \operatorname{rank}(\mathbf{Z}_1) \leqslant q$ *and* $\operatorname{rank}(\mathbf{Z}_2) = q$.
B. $\mathbf{N} = \mathbf{K}_1\mathbf{Z}_1^+$, *if* $\operatorname{rank}(\mathbf{Z}_1) > 0$ *and* $\mathbf{Z}_2 = \mathbf{0}$ *and* $\mathbf{Y} = \mathbf{0}$.
C. $\mathbf{N} = \mathbf{Z}_2^{+*}\mathbf{K}_2$, *if* $\mathbf{Z}_1 = \mathbf{0}$ *and* $\operatorname{rank}(\mathbf{Z}_2) > 0$ *and* $\mathbf{Y} = \mathbf{0}$.
D. $\mathbf{N} = \mathbf{K}_1\mathbf{Y}^+\mathbf{K}_2$, *if* $\mathbf{Z}_1 = \mathbf{0}$, $\mathbf{Z}_2 = \mathbf{0}$ *and* $0 \leqslant \operatorname{rank}(\mathbf{Y}) \leqslant q$.

Proof. The relaxation of full rank requirements on $\mathbf{Z}_1$, $\mathbf{Z}_2$, and $\mathbf{Y}$ means that identities (e) and (k) no longer hold and $\mathbf{Y}^{-1}$ no longer exists. These identities are not required though as we only need to verify Moore–Penrose condition (I) and these identities are only needed for (III) and (IV). □

Remark 2

(1) Part D also holds if $\mathbf{Z}_1 = \mathbf{0}$ and $\mathbf{W}_1\mathbf{Q}_{\mathbf{Y}}\mathbf{W}_2^* = \mathbf{0}$ or $\mathbf{Z}_2 = \mathbf{0}$ and $\mathbf{W}_1\mathbf{Q}_{\mathbf{Y}^*}\mathbf{W}_2^* = \mathbf{0}$.
(2) The conditions of Theorem 1.F can be relaxed to $\mathbf{M}^{-} = \mathbf{A}^+ - (\mathbf{S}_1\mathbf{Z}_1 + \mathbf{K}_1)(\mathbf{Y} + \mathbf{W}_2^*\mathbf{S}_1\mathbf{Z}_1)^{+}\mathbf{T}_1 + \mathbf{S}_1$, with $\mathbf{S}_1 = \mathbf{A}^+\mathbf{K}_2^*\mathbf{Y}^{+*}\mathbf{Z}_1^*$, and $\mathbf{T}_1 = \mathbf{W}_2^*\mathbf{S}_1 + \mathbf{K}_2$, if $\mathbf{Z}_2 = \mathbf{0}$, $0 \leqslant \operatorname{rank}(\mathbf{Y}) \leqslant q$ and $(\mathbf{A}\mathbf{S}_1 - \mathbf{I}_n)\mathbf{Z}_1\mathbf{Q}_{\mathbf{R}_1^*}\mathbf{W}_2^* = \mathbf{0}$, where $\mathbf{R}_1 = \mathbf{Y} + \mathbf{W}_2^*\mathbf{S}_1\mathbf{Z}_1$.
(3) The conditions of Theorem 1.G can be relaxed to $\mathbf{M}^{-} = \mathbf{A}^+ - \mathbf{T}_2(\mathbf{Y} + \mathbf{Z}_2^*\mathbf{S}_2\mathbf{W}_1)^{+}(\mathbf{Z}_2^*\mathbf{S}_2 + \mathbf{K}_2) + \mathbf{S}_2$, with $\mathbf{S}_2 = \mathbf{Z}_2\mathbf{Y}^{+*}\mathbf{K}_1^*\mathbf{A}^+$, and $\mathbf{T}_2 = \mathbf{S}_2\mathbf{W}_1 + \mathbf{K}_1$, if $\mathbf{Z}_1 = \mathbf{0}$, $0 \leqslant \operatorname{rank}(\mathbf{Y}) \leqslant q$, and $\mathbf{W}_1\mathbf{Q}_{\mathbf{R}_2}\mathbf{Z}_2^*(\mathbf{S}_2\mathbf{A} - \mathbf{I}_m) = \mathbf{0}$, where $\mathbf{R}_2 = \mathbf{Y} + \mathbf{Z}_2^*\mathbf{S}_2\mathbf{W}_1$.

3. Results for (quasi-)Kronecker structures

We now turn to the question of Moore–Penrose inverses for matrices of the form (2), i.e., $\mathbf{A} = \operatorname{bdiag}(\mathbf{A}_k)$, $\mathbf{B} = \mathbf{u}\mathbf{v}^* \otimes \mathbf{E}$, which we write in a form such that we can apply Theorem 1. Assume that $0 < \operatorname{rank}(\mathbf{E}) = q \leqslant m$ and write $\mathbf{E} = \tilde{\mathbf{U}}\tilde{\mathbf{D}}\tilde{\mathbf{V}}^*$ where the diagonal entries of $\tilde{\mathbf{D}}$ are sorted in decreasing order. Let $\tilde{\mathbf{U}}_{1:q}$ and $\tilde{\mathbf{V}}_{1:q}$ be the first q columns of $\tilde{\mathbf{U}}$ and $\tilde{\mathbf{V}}$ and let $\tilde{\mathbf{D}}_{1:q}^{1/2}$ be the diagonal square matrix with entries $\tilde{\mathbf{D}}_{ii}^{1/2}$, $i = 1, \ldots, q$. Finally, let $\mathbf{U} = \tilde{\mathbf{U}}_{1:q}\tilde{\mathbf{D}}_{1:q}^{1/2}$ and $\mathbf{V} = \tilde{\mathbf{V}}_{1:q}\tilde{\mathbf{D}}_{1:q}^{1/2}$, i.e., $\mathbf{E} = \mathbf{U}\mathbf{V}^*$. We now decompose $\mathbf{u} \otimes \mathbf{U}$ and $\mathbf{v} \otimes \mathbf{V}$ into $\mathbf{W}_1 + \mathbf{Z}_1$ and $\mathbf{W}_2 + \mathbf{Z}_2$ by defining $\mathbf{W}_1 = \mathbf{P}_{\mathbf{A}}(\mathbf{u} \otimes \mathbf{U})$, $\mathbf{Z}_1 = \mathbf{Q}_{\mathbf{A}}(\mathbf{u} \otimes \mathbf{U})$, $\mathbf{W}_2 = \mathbf{P}_{\mathbf{A}^*}(\mathbf{v} \otimes \mathbf{V})$ and $\mathbf{Z}_2 = \mathbf{Q}_{\mathbf{A}^*}(\mathbf{v} \otimes \mathbf{V})$.

Corollary 4. *Let* $\mathbf{A}_1, \ldots, \mathbf{A}_p$, $\mathbf{E} = \mathbf{U}\mathbf{V}^*$ *be arbitrary* $n \times m$ *complex matrices and let* $\mathbf{u} = (u_1, \ldots, u_p)^T$ *and* $\mathbf{v} = (v_1, \ldots, v_p)^T$ *be complex p-vectors such that* $\mathbf{u}^*\mathbf{u} = \mathbf{v}^*\mathbf{v} = 1$. *Let* $\mathbf{H}_1 = \mathbf{U}^*(\mathbf{I}_n - \sum_k |u_k|^2\mathbf{A}_k\mathbf{A}_k^+)\mathbf{U}$, $\mathbf{H}_2 = \mathbf{V}^*(\mathbf{I}_m - \sum_k |v_k|^2\mathbf{A}_k^+\mathbf{A}_k)\mathbf{V}$, $\operatorname{brow}(\mathbf{G}_1^j) = (\operatorname{bcol}(u_j\mathbf{A}_j^+\mathbf{U}))^+$, $\operatorname{bcol}(\mathbf{G}_2^i) = (\operatorname{brow}(\bar{v}_i\mathbf{V}^*\mathbf{A}_i^+))^+$ *and* $\mathbf{Y} = \mathbf{I}_q + \mathbf{V}^*\sum_k u_k\bar{v}_k\mathbf{A}_k^+\mathbf{U}$. *Under the assumptions detailed below, the Moore–Penrose inverse of* $\mathbf{M} = \operatorname{bdiag}(\mathbf{A}_k) + \mathbf{u}\mathbf{v}^* \otimes \mathbf{U}\mathbf{V}^*$ *is*

$$\mathbf{M}^+ = \operatorname{bdiag}(\mathbf{A}_k^+) - \operatorname{block}(\mathbf{N}_{ij}),$$

where

A. $\mathbf{N}_{ij} = v_i\bar{v}_j\mathbf{Q}_{\mathbf{A}_i^*}\mathbf{V}\mathbf{H}_2^{-1}\mathbf{V}^*\mathbf{A}_j^+ + u_i\bar{u}_j\mathbf{A}_i^+\mathbf{U}\mathbf{H}_1^{-1}\mathbf{U}^*\mathbf{Q}_{\mathbf{A}_j} - v_i\bar{u}_j\mathbf{Q}_{\mathbf{A}_i^*}\mathbf{V}\mathbf{H}_2^{-1}\mathbf{Y}\mathbf{H}_1^{-1}\mathbf{U}^*\mathbf{Q}_{\mathbf{A}_j}$, *if there exists an* r *and an* s, *such that* $\operatorname{rank}(u_r\mathbf{Q}_{\mathbf{A}_r}\mathbf{U}) = q$ *and* $\operatorname{rank}(v_s\mathbf{Q}_{\mathbf{A}_s^*}\mathbf{V}) = q$.

B. $\mathbf{N}_{ij} = u_i\bar{u}_j\mathbf{A}_i^+\mathbf{U}\mathbf{H}_1^{-1}\mathbf{U}^*\mathbf{Q}_{\mathbf{A}_j} + \bar{v}_j\mathbf{A}_i^+\mathbf{G}_2^i\mathbf{V}^*\mathbf{A}_j^+$, *if there exists an* r, *such that* $\operatorname{rank}(u_r\mathbf{Q}_{\mathbf{A}_r}\mathbf{U}) = q$, $\mathbf{Z}_2 = \mathbf{0}$ *and* $\mathbf{Y} = \mathbf{0}$.

C. $\mathbf{N}_{ij} = v_i\bar{v}_j\mathbf{Q}_{\mathbf{A}_i^*}\mathbf{V}\mathbf{H}_2^{-1}\mathbf{V}^*\mathbf{A}_j^+ + u_i\mathbf{A}_i^+\mathbf{U}\mathbf{G}_1^j\mathbf{A}_j^+$, *if* $\mathbf{Z}_1 = \mathbf{0}$, *if there exists an* s, *such that* $\operatorname{rank}(v_s\mathbf{Q}_{\mathbf{A}_s^*}\mathbf{V}) = q$ *and* $\mathbf{Y} = \mathbf{0}$.

D. $\mathbf{N}_{ij} = u_i\mathbf{A}_i^+\mathbf{U}\mathbf{G}_1^j\mathbf{A}_j^+ + \bar{v}_j\mathbf{A}_i^+\mathbf{G}_2^i\mathbf{V}^*\mathbf{A}_j^+ - u_i\bar{v}_j\mathbf{A}_i^+\mathbf{U}\mathbf{G}_{12}\mathbf{V}^*\mathbf{A}_j^+$ *with* $\mathbf{G}_{12} = \sum_k \mathbf{G}_1^k\mathbf{A}_k^+\mathbf{G}_2^k$, *if* $\mathbf{Z}_1 = \mathbf{0}$, $\mathbf{Z}_2 = \mathbf{0}$ *and* $\mathbf{Y} = \mathbf{0}$.

E. $\mathbf{N}_{ij} = u_i\bar{v}_j\mathbf{A}_i^+\mathbf{U}\mathbf{Y}^{-1}\mathbf{V}^*\mathbf{A}_j^+$, *if* $\mathbf{Z}_1 = \mathbf{0}$, $\mathbf{Z}_2 = \mathbf{0}$ *and* $\operatorname{rank}(\mathbf{Y}) = q$.

F. $\mathbf{N}_{ij} = (u_i\mathbf{A}_i^+\mathbf{U} + \mathbf{S}_{1,i}\mathbf{H}_1)(\mathbf{Y} + \mathbf{V}^*\sum_k \bar{v}_k\mathbf{S}_{1,k}\mathbf{H}_1)^{-1}\mathbf{T}_{1,j} - \bar{u}_j\mathbf{S}_{1,i}\mathbf{U}^*\mathbf{Q}_{\mathbf{A}_j}$, *with* $\mathbf{S}_{1,i} = v_i\mathbf{A}_i^{+}\mathbf{A}_i^{+*}\mathbf{V}\mathbf{Y}^{*-1}$ *and* $\mathbf{T}_{1,j} = \bar{v}_j\mathbf{V}^*\mathbf{A}_j^+ + \bar{u}_j\mathbf{V}^*\sum_k \bar{v}_k\mathbf{S}_{1,k}\mathbf{U}^*\mathbf{Q}_{\mathbf{A}_j}$, *if* $\mathbf{Z}_2 = \mathbf{0}$ *and* $\operatorname{rank}(\mathbf{Y}) = q$.

G. $\mathbf{N}_{ij} = \mathbf{T}_{2,i}(\mathbf{Y} + \mathbf{H}_2\sum_k \mathbf{u}_k\mathbf{S}_{2,k}\mathbf{U})^{-1}(\bar{v}_j\mathbf{V}^*\mathbf{A}_j^+ + \mathbf{H}_2\mathbf{S}_{2,j}) - v_i\mathbf{Q}_{\mathbf{A}_i^*}\mathbf{V}\mathbf{S}_{2,j}$, *with* $\mathbf{S}_{2,j} = \bar{u}_j\mathbf{Y}^{*-1}\mathbf{U}^*\mathbf{A}_j^{+*}\mathbf{A}_j^+$, *and* $\mathbf{T}_{2,i} = u_i\mathbf{A}_i^+\mathbf{U} + v_i\mathbf{Q}_{\mathbf{A}_i^*}\mathbf{V}\sum_k u_k\mathbf{S}_{2,k}\mathbf{U}$, *if* $\mathbf{Z}_1 = \mathbf{0}$ *and* $\operatorname{rank}(\mathbf{Y}) = q$.

Proof. By construction, $\mathbf{u}\otimes\mathbf{U} = \mathbf{W}_1 + \mathbf{Z}_1$, and the columns of $\mathbf{W}_1$ and $\mathbf{Z}_1$ are contained in $R(\mathbf{A})$ and $R(\mathbf{A})^\perp$, respectively. Similarly, the columns of $\mathbf{W}_2$ and $\mathbf{Z}_2$ are contained in $R(\mathbf{A}^*)$ and $R(\mathbf{A}^*)^\perp$, respectively. We use the block structure of the matrices and have $\mathbf{W}_1 = \operatorname{bcol}(u_i\mathbf{P}_{\mathbf{A}_i}\mathbf{U})$, $\mathbf{W}_2 = \operatorname{bcol}(v_i\mathbf{P}_{\mathbf{A}_i^*}\mathbf{V})$, $\mathbf{Z}_1 = \operatorname{bcol}(u_i\mathbf{Q}_{\mathbf{A}_i}\mathbf{U})$ and $\mathbf{Z}_2 = \operatorname{bcol}(v_i\mathbf{Q}_{\mathbf{A}_i^*}\mathbf{V})$. For all i, $\operatorname{rank}(u_i\mathbf{Q}_{\mathbf{A}_i}\mathbf{U}) \leqslant \operatorname{rank}(\mathbf{Z}_1)$, and, similarly, $\operatorname{rank}(v_i\mathbf{Q}_{\mathbf{A}_i^*}\mathbf{V}) \leqslant \operatorname{rank}(\mathbf{Z}_2)$. If $\operatorname{rank}(u_r\mathbf{Q}_{\mathbf{A}_r}\mathbf{U}) = q$ and $\operatorname{rank}(v_s\mathbf{Q}_{\mathbf{A}_s^*}\mathbf{V}) = q$, the inverses of $\mathbf{H}_1$ and $\mathbf{H}_2$, respectively, exist. Notice that $\mathbf{Z}_2^{+*} = \mathbf{Z}_2(\mathbf{Z}_2^*\mathbf{Z}_2)^+$. Thus for the special case of quasi-Kronecker form, $\mathbf{Z}_2^{+*}$ becomes $\operatorname{bcol}(v_i\mathbf{Q}_{\mathbf{A}_i^*}\mathbf{V})\mathbf{H}_2^{-1}$. Similarly, $\mathbf{Z}_1^+$ becomes $\mathbf{H}_1^{-1}\operatorname{brow}(\bar{u}_i\mathbf{U}^*\mathbf{Q}_{\mathbf{A}_i})$.

A. Given that there exists an r and an s, such that $\operatorname{rank}(u_r\mathbf{Q}_{\mathbf{A}_r}\mathbf{U}) = q$ and $\operatorname{rank}(v_s\mathbf{Q}_{\mathbf{A}_s^*}\mathbf{V}) = q$, we have that $\operatorname{rank}(\mathbf{Z}_1) = q$ and $\operatorname{rank}(\mathbf{Z}_2) = q$ and we can use Theorem 1. From Theorem 1. A we have,

$$
\begin{aligned}
\mathbf{N} &= \mathbf{Z}_2^{+*}\mathbf{W}_2^*\mathbf{A}^+ + \mathbf{A}^+\mathbf{W}_1\mathbf{Z}_1^+ - \mathbf{Z}_2^{+*}\mathbf{Y}\mathbf{Z}_1^+\\
&= \operatorname{bcol}(v_i\mathbf{Q}_{\mathbf{A}_i^*}\mathbf{V})\mathbf{H}_2^{-1}\operatorname{brow}(\bar{v}_j\mathbf{V}^*\mathbf{P}_{\mathbf{A}_j^*})\operatorname{bdiag}(\mathbf{A}_k^+)\\
&\quad + \operatorname{bdiag}(\mathbf{A}_k^+)\operatorname{bcol}(u_i\mathbf{P}_{A_i}\mathbf{U})\mathbf{H}_1^{-1}\operatorname{brow}(\bar{u}_j\mathbf{U}^*\mathbf{Q}_{\mathbf{A}_j})\\
&\quad - \operatorname{bcol}(v_i\mathbf{Q}_{\mathbf{A}_i^*}\mathbf{V})\mathbf{H}_2^{-1}\mathbf{Y}\mathbf{H}_1^{-1}\operatorname{brow}(\bar{u}_j\mathbf{U}^*\mathbf{Q}_{\mathbf{A}_j})\\
&= \operatorname{bcol}(v_i\mathbf{Q}_{\mathbf{A}_i^*}\mathbf{V})\mathbf{H}_2^{-1}\operatorname{brow}(\bar{v}_j\mathbf{V}^*\mathbf{A}_j^+)\\
&\quad + \operatorname{bcol}(u_i\mathbf{A}_i^+\mathbf{U})\mathbf{H}_1^{-1}\operatorname{brow}(\bar{u}_j\mathbf{U}^*\mathbf{Q}_{\mathbf{A}_j})\\
&\quad - \operatorname{block}(v_i\bar{u}_j\mathbf{Q}_{\mathbf{A}_i^*}\mathbf{V}\mathbf{H}_2^{-1}\mathbf{Y}\mathbf{H}_1^{-1}\mathbf{U}^*\mathbf{Q}_{\mathbf{A}_j})\\
&= \operatorname{block}(v_i\bar{v}_j\mathbf{Q}_{\mathbf{A}_i^*}\mathbf{V}\mathbf{H}_2^{-1}\mathbf{V}^*\mathbf{A}_j^+ + u_i\bar{u}_j\mathbf{A}_i^+\mathbf{U}\mathbf{H}_1^{-1}\mathbf{U}^*\mathbf{Q}_{\mathbf{A}_j}\\
&\quad - v_i\bar{u}_j\mathbf{Q}_{\mathbf{A}_i^*}\mathbf{V}\mathbf{H}_2^{-1}\mathbf{Y}\mathbf{H}_1^{-1}\mathbf{U}^*\mathbf{Q}_{\mathbf{A}_j}) = \operatorname{block}(\mathbf{N}_{ij}).
\end{aligned}
$$

The remaining cases B–G can also be rewritten in a straightforward manner in terms of $\mathbf{A}_k$, $\mathbf{U}$, $\mathbf{V}$, $\mathbf{u}$, $\mathbf{v}$, etc. using the results of Theorem 1.B–G. □

Similar to Theorem 1, the rank q assumptions can be relaxed and in the case of symmetric matrices, part B and part C are irrelevant since this implies $\operatorname{rank}(\mathbf{Z}_1) = \operatorname{rank}(\mathbf{Z}_2)$. In the terms containing $\mathbf{H}_\ell$ in parts A–C, many of the diagonal matrices $\tilde{\mathbf{D}}_{1:q}^{1/2}$ disappear. Eq. (3) is a particular case of part E.

If we assume that $\mathbf{A} = \mathbf{I}_p\otimes\mathbf{A}_1$, further simplifications are possible, for example $\mathbf{H}_1 = \mathbf{U}^*\mathbf{Q}_{\mathbf{A}_1}\mathbf{U}$, $\mathbf{H}_2 = \mathbf{V}^*\mathbf{Q}_{\mathbf{A}_1^*}\mathbf{V}$ and $\mathbf{Y} = \mathbf{I}_q + \mathbf{v}^*\mathbf{u}\mathbf{V}^*\mathbf{A}_1^+\mathbf{U}$. The matrix $\operatorname{block}(\mathbf{N}_{ij})$ given in Corollary 4.A–4.G simplify, respectively, to (the index on the $\mathbf{A}$ matrix has been dropped for ease of notation)

$$
\begin{aligned}
\operatorname{block}(\mathbf{N}_{ij}) &= \mathbf{u}\mathbf{u}^* \otimes \mathbf{A}^+\mathbf{U}\mathbf{H}_1^{-1}\mathbf{U}^*\mathbf{Q}_{\mathbf{A}} + \mathbf{v}\mathbf{v}^* \otimes \mathbf{Q}_{\mathbf{A}^*}\mathbf{V}\mathbf{H}_2^{-1}\mathbf{V}^*\mathbf{A}^+ \\
&\quad - \mathbf{v}\mathbf{u}^* \otimes \mathbf{Q}_{\mathbf{A}^*}\mathbf{V}\mathbf{H}_2^{-1}\mathbf{Y}\mathbf{H}_1^{-1}\mathbf{U}^*\mathbf{Q}_{\mathbf{A}}, \\
\operatorname{block}(\mathbf{N}_{ij}) &= \mathbf{v}\mathbf{v}^* \otimes \mathbf{Q}_{\mathbf{A}^*}\mathbf{V}\mathbf{H}_2^{-1}\mathbf{V}^*\mathbf{A}^+ + \mathbf{u}\mathbf{u}^* \otimes \mathbf{A}^+\mathbf{U}(\mathbf{A}^+\mathbf{U})^+\mathbf{A}^+, \\
\operatorname{block}(\mathbf{N}_{ij}) &= \mathbf{u}\mathbf{u}^* \otimes \mathbf{A}^+\mathbf{U}\mathbf{H}_1^{-1}\mathbf{U}^*\mathbf{Q}_{\mathbf{A}} + \mathbf{v}\mathbf{v}^* \otimes \mathbf{A}^+(\mathbf{V}^*\mathbf{A}^+)^+\mathbf{V}^*\mathbf{A}^+, \\
\operatorname{block}(\mathbf{N}_{ij}) &= \mathbf{u}\mathbf{u}^* \otimes \mathbf{A}^+\mathbf{U}(\mathbf{A}^+\mathbf{U})^+\mathbf{A}^+ + \mathbf{v}\mathbf{v}^* \otimes \mathbf{A}^+(\mathbf{V}^*\mathbf{A}^+)^+\mathbf{V}^*\mathbf{A}^+ \\
&\quad - \mathbf{u}\mathbf{u}^*\mathbf{v}\mathbf{v}^* \otimes \mathbf{A}^+\mathbf{U}(\mathbf{A}^+\mathbf{U})^+\mathbf{A}^+(\mathbf{V}^*\mathbf{A}^+)^+\mathbf{V}^*\mathbf{A}^+, \\
\operatorname{block}(\mathbf{N}_{ij}) &= \mathbf{u}\mathbf{v}^* \otimes \mathbf{A}^+\mathbf{U}\mathbf{Y}^{-1}\mathbf{V}^*\mathbf{A}^+, \\
\operatorname{block}(\mathbf{N}_{ij}) &= \mathbf{v}\mathbf{u}^* \otimes \mathbf{A}^+\mathbf{A}^{+*}\mathbf{V}\mathbf{Y}^{*-1}(\mathbf{I}_q + \mathbf{H}_1\mathbf{R}_1^{-1}\mathbf{V}^*\mathbf{A}^+\mathbf{A}^{+*}\mathbf{V}\mathbf{Y}^{*-1})\mathbf{U}^*\mathbf{Q}_{\mathbf{A}} \\
&\quad + \mathbf{u}\mathbf{v}^* \otimes \mathbf{A}^+\mathbf{U}\mathbf{R}_1^{-1}\mathbf{V}^*\mathbf{A}^+ + \mathbf{v}\mathbf{v}^* \otimes \mathbf{A}^+\mathbf{A}^{+*}\mathbf{V}\mathbf{Y}^{*-1}\mathbf{H}_1\mathbf{R}_1^{-1}\mathbf{V}^*\mathbf{A}^+ \\
&\quad + \mathbf{u}\mathbf{u}^* \otimes \mathbf{A}^+\mathbf{U}\mathbf{R}_1^{-1}\mathbf{V}^*\mathbf{A}^+\mathbf{A}^{+*}\mathbf{V}\mathbf{Y}^{*-1}\mathbf{U}^*\mathbf{Q}_{\mathbf{A}}, \\
\operatorname{block}(\mathbf{N}_{ij}) &= \mathbf{v}\mathbf{u}^* \otimes \mathbf{Q}_{\mathbf{A}^*}\mathbf{V}\mathbf{Y}^{*-1}(\mathbf{I}_q + \mathbf{U}^*\mathbf{A}^{+*}\mathbf{A}^+\mathbf{U}\mathbf{R}_2^{-1}\mathbf{H}_2\mathbf{Y}^{*-1})\mathbf{U}^*\mathbf{A}^{+*}\mathbf{A}^+ \\
&\quad + \mathbf{u}\mathbf{v}^* \otimes \mathbf{A}^+\mathbf{U}\mathbf{R}_2^{-1}\mathbf{V}^*\mathbf{A}^+ + \mathbf{u}\mathbf{u}^* \otimes \mathbf{A}^+\mathbf{U}\mathbf{R}_2^{-1}\mathbf{H}_2\mathbf{Y}^{*-1}\mathbf{U}^*\mathbf{A}^{+*}\mathbf{A}^+ \\
&\quad + \mathbf{v}\mathbf{v}^* \otimes \mathbf{Q}_{\mathbf{A}^*}\mathbf{V}\mathbf{Y}^{*-1}\mathbf{U}^*\mathbf{A}^{+*}\mathbf{A}^+\mathbf{U}\mathbf{R}_2^{-1}\mathbf{V}^*\mathbf{A}^+,
\end{aligned}
$$

where $\mathbf{R}_1 = \mathbf{Y} + \mathbf{V}^*\mathbf{A}^+\mathbf{A}^{+*}\mathbf{V}\mathbf{Y}^{*-1}\mathbf{H}_1$ and $\mathbf{R}_2 = \mathbf{Y} + \mathbf{H}_2\mathbf{Y}^{*-1}\mathbf{U}^*\mathbf{A}^{+*}\mathbf{A}^+\mathbf{U}$.

4. Application to hierarchical multivariate models

We now turn to matrices that occur in statistical models, i.e., the matrices are symmetric and positive definite or positive semi-definite. The research presented here was motivated by the following problem. Assume that $\mathcal{Y}$ is a Gaussian random n-vector, with mean $\boldsymbol{\mu}$ and covariance $\mathbf{C}$, denoted as $\mathcal{Y} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$ and let $\mathcal{X}_k = c_k\mathcal{Y} + \boldsymbol{\varepsilon}_k$, $k = 1, \ldots, p$, with $\boldsymbol{\varepsilon}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{S}_k)$, independent of each other and of $\mathcal{Y}$. The joint distribution of $\mathcal{X}_1, \ldots, \mathcal{X}_p$ is again Gaussian and has covariance structure $\boldsymbol{\Omega} = \operatorname{bdiag}(\mathbf{S}_k) + \mathbf{c}\mathbf{c}^T \otimes \mathbf{C}$, where $\mathbf{c} = (c_1, \ldots, c_p)^T$. In spatial or spatio-temporal applications, often only one or a few observations are available and it is not possible to estimate the mean vector and all covariance matrices consistently from the data. Instead, the mean vector and covariance matrices are parameterized by a low dimensional parameter [17], e.g. $\boldsymbol{\mu} = \boldsymbol{\mu}(\boldsymbol{\theta}_\mu)$, $\mathbf{S}_k = \mathbf{S}_k(\boldsymbol{\theta}_k)$, $\mathbf{C} = \mathbf{C}(\boldsymbol{\theta}_{\mathbf{C}})$. Given data or observations of $\mathcal{X}_1, \ldots \mathcal{X}_p$, maximum likelihood estimation of $\boldsymbol{\theta} = (\boldsymbol{\theta}_\mu^T, \boldsymbol{\theta}_1^T, \ldots, \boldsymbol{\theta}_p^T, \boldsymbol{\theta}_{\mathbf{C}}^T)^T$ involves many evaluations of $\det(\boldsymbol{\Omega})$ and quadratic forms in $\boldsymbol{\Omega}^{-1}$. Similarly, best linear unbiased predictors (BLUPs) of $\mathcal{Y}$ or of unobserved components of $\mathcal{X}_k$ require solving linear systems based on (conditional) covariance matrices. While p is usually of moderate size in many statistical applications, n is often large or very large. To circumvent working with the entire matrix $\boldsymbol{\Omega}$, researchers often use simpler covariance models [18], sparse matrix algebra [19,20], or lower dimensional approximations [21,22]. Hence, there is a clear need for closed form expressions of the (Moore–Penrose) inverse and the determinant of $\boldsymbol{\Omega}$ in terms of $\mathbf{S}_k$ and $\mathbf{C}$.

Let $\mathbf{S}_1, \ldots, \mathbf{S}_p, \mathbf{C}$ be symmetric positive definite $n \times n$ matrices and let $\mathbf{c}$ be a p-vector and (without loss of generality) $\mathbf{c}^*\mathbf{c} = 1$. Corollary 4.E (or working in this simple case directly with the Woodbury formula (3)) yields

$$
\boldsymbol{\Omega}^{-1} = \operatorname{bdiag}(\mathbf{S}_k^{-1}) - \operatorname{block}\left(c_i c_j \mathbf{S}_i^{-1} \left(\mathbf{C}^{-1} + \sum_{k=1}^{p} c_k^2 \mathbf{S}_k^{-1} \right)^{-1} \mathbf{S}_j^{-1} \right). \tag{4}
$$

To calculate the determinant of $\boldsymbol{\Omega}$, we use the decomposition

$$
\begin{pmatrix} \mathbf{I}_n & \mathbf{0} \\ \mathbf{V}^* & \mathbf{I}_q \end{pmatrix} \begin{pmatrix} \mathbf{I}_n + \mathbf{U}\mathbf{V}^* & \mathbf{U} \\ \mathbf{0} & \mathbf{I}_q \end{pmatrix} \begin{pmatrix} \mathbf{I}_n & \mathbf{0} \\ -\mathbf{V}^* & \mathbf{I}_q \end{pmatrix} = \begin{pmatrix} \mathbf{I}_n & \mathbf{U} \\ \mathbf{0} & \mathbf{I}_q + \mathbf{V}^*\mathbf{U} \end{pmatrix}
$$

and by setting $\mathbf{U} = \text{bdiag}(\mathbf{S}_k^{-1})(\mathbf{c} \otimes \mathbf{I}_n)$ and $\mathbf{V} = \mathbf{c} \otimes \mathbf{C}$ we have

$$\begin{aligned}\det(\boldsymbol{\Omega}) &= \det\left(\text{bdiag}(\mathbf{S}_k)\right) \det\left(\mathbf{I}_{np} + \text{bdiag}(\mathbf{S}_k^{-1})(\mathbf{c}\mathbf{c}^T \otimes \mathbf{C})\right) \\ &= \prod_{i=1}^{p} \det(\mathbf{S}_i) \det\left(\mathbf{I}_n + \mathbf{C}\sum_{k=1}^{p} c_k^2 \mathbf{S}_k^{-1}\right).\end{aligned}$$

Consider the joint distribution of $\mathcal{X}_1, \ldots, \mathcal{X}_p, \mathcal{Y}$, let $\boldsymbol{\Omega}_0 = \text{bdiag}(\mathbf{S}_1, \ldots, \mathbf{S}_p, \mathbf{0}) + \mathbf{c}_0\mathbf{c}_0^T \otimes 2\mathbf{C}$, $\mathbf{c}_0 = (\mathbf{c}^T, 1)^T/\sqrt{2}$ with $\mathbf{S}_k$ and $\mathbf{C}$ symmetric positive definite n matrices. Then, by Corollary 4.A, $\boldsymbol{\Omega}_0^{-1} = \text{block}(\boldsymbol{\Omega}_0^{ij})$, where

$$\boldsymbol{\Omega}_0^{ij} = \begin{cases} \mathbf{S}_i^{-1} & i = j,\ 1 \leqslant i \leqslant p, \\ \mathbf{0} & i \neq j,\ 1 \leqslant i, j \leqslant p, \\ -c_i\mathbf{S}_i^{-1} & j = p+1,\ 1 \leqslant i \leqslant p, \\ -c_j\mathbf{S}_j^{-1} & i = p+1,\ 1 \leqslant j \leqslant p, \\ \mathbf{C}^{-1} + \sum_{k=1}^{p} c_k^2\mathbf{S}_k^{-1} & i = j = p+1. \end{cases} \tag{5}$$

Symmetric positive semi-definite covariance matrices occur, for example, in intrinsic Gaussian random fields and are often used as priors, e.g. [23]. If we write $\mathbf{C} = \boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}^T$, with $\boldsymbol{\Gamma}$ an $n \times q$ matrix with orthonormal column vectors, the blocks $\boldsymbol{\Omega}_0^{ij}$ are $\mathbf{S}_i^{-1}$, $\mathbf{0}$, $-c_i\mathbf{S}_i^{-1}\boldsymbol{\Gamma}\boldsymbol{\Gamma}^T$, $-c_j\boldsymbol{\Gamma}\boldsymbol{\Gamma}^T\mathbf{S}_j^{-1}$, and $\mathbf{C}^+ + \sum_k c_k^2\boldsymbol{\Gamma}\boldsymbol{\Gamma}^T\mathbf{S}_k^{-1}\boldsymbol{\Gamma}\boldsymbol{\Gamma}^T$ with a similar structure as in (5). Other cases can be derived similarly from Corollary 4.

Another matrix of interest is the conditional covariance of the set $\{\mathcal{X}_k\}$, $k \in S \subset \{1, \ldots, p\}$ given the set $\{\mathcal{X}_\ell\}$, $\ell \notin S$. Without loss of generality, we assume that $S = \{1, \ldots, r\}$ and write $\mathcal{X}_S = \{\mathcal{X}_1 \ \ldots, \mathcal{X}_r\}$ and $\mathcal{X}_{S^c} = \{\mathcal{X}_{r+1} \ \ldots, \mathcal{X}_p\}$. Let $\mathbf{c} = (\mathbf{c}_1^T, \mathbf{c}_2^T)^T$, i.e., $\mathbf{c}_1$ contains the first r components of $\mathbf{c}$. The conditional variance is given by

$$\begin{aligned}\text{Var}(\mathcal{X}_S|\mathcal{X}_{S^c}) &= \text{Var}(\mathcal{X}_S) - \text{Cov}(\mathcal{X}_S, \mathcal{X}_{S^c})\,\text{Var}(\mathcal{X}_{S^c})^{-1}\,\text{Cov}(\mathcal{X}_{S^c}, \mathcal{X}_S) \\ &= \text{bdiag}(\mathbf{S}_1, \ldots, \mathbf{S}_r) + \mathbf{c}_1\mathbf{c}_1^T \otimes \mathbf{C} - (\mathbf{c}_1\mathbf{c}_2^T \otimes \mathbf{C}) \\ &\quad \times \left(\text{bdiag}(\mathbf{S}_{r+1}, \ldots, \mathbf{S}_p) + \mathbf{c}_2\mathbf{c}_2^T \otimes \mathbf{C}\right)^{-1} (\mathbf{c}_2\mathbf{c}_1^T \otimes \mathbf{C}) \\ &= \text{bdiag}(\mathbf{S}_1, \ldots, \mathbf{S}_r) + \mathbf{c}_1\mathbf{c}_1^T \otimes \left(\mathbf{C}^{-1} + \sum_{k=r+1}^{p} c_k^2\mathbf{S}_k^{-1}\right)^{-1},\end{aligned} \tag{6}$$

where we used (4) and the identity

$(\mathbf{C}^{-1} + \mathbf{G})^{-1} = (\mathbf{I} - \mathbf{C}\mathbf{G}(\mathbf{I} + \mathbf{C}\mathbf{G})^{-1})\mathbf{C} = \mathbf{C} - \mathbf{C}\mathbf{G}(\mathbf{I} - (\mathbf{C}^{-1} + \mathbf{G})^{-1}\mathbf{G})\mathbf{C}$, for appropriate matrices $\mathbf{G}$. As a special case of (6) we have

$$\text{Var}(\mathcal{Y}|\mathcal{X}_1, \ldots, \mathcal{X}_p) = \left(\mathbf{C}^{-1} + \sum_{k=1}^{p} c_k^2\mathbf{S}_k^{-1}\right)^{-1}.$$

References

[1] K. Miller, On the inverse of the sum of matrices, Math. Mag. 54 (1981) 67–72.
[2] H. Henderson, S. Searle, On deriving the inverse of a sum of matrices, SIAM Rev. 23 (1981) 53–60.
[3] W. Hager, Updating the inverse of a matrix, SIAM Rev. 31 (1989) 221–239.
[4] C. Meyer Jr., Generalized inversion of modified matrices, SIAM J. Appl. Math. 24 (1973) 315–323.
[5] J. Baksalary, O.M. Baksalary, G. Trenkler, A revisitation of formulae for the Moore–Penrose inverse of modified matrices, Linear Algebra Appl. 372 (2003) 207–224.

[6] K. Riedel, A Sherman–Morrison–Woodbury identity for rank augmenting matrices with application to centering, SIAM J. Matrix Anal. Appl. 13 (1992) 659–662.
[7] J. Fill, D. Fishkind, The Moore–Penrose generalized inverse for sums of matrices, SIAM J. Matrix Anal. Appl. 21 (2000) 629–635.
[8] S. Lai, B. Vemuri, Generalized capacitance matrix theorems for solving linear systems, SIAM J. Sci. Comput. 19 (1998) 1024–1045.
[9] T. Steerneman, F. van Perlo-ten Kleij, Properties of the matrix $A - XY^*$, Linear Algebra Appl. 410 (2005) 70–86.
[10] C. Rohde, Generalized inverses of partitioned matrices, J. Soc. Indust. Appl. Math. 13 (1965) 1033–1035.
[11] C. Meyer Jr., Generalized inverses and ranks of block matrices, SIAM J. Appl. Math. 25 (1973) 597–602.
[12] C. Meyer Jr., Generalized inverses of block triangular matrices, SIAM J. Appl. Math. 19 (1970) 741–750.
[13] D. Harville, Matrix Algebra from a Statistician's Perspective, Springer, Heidelberg, 1997.
[14] J. Miao, General expressions for the Moore–Penrose inverse of a 2×2 block matrix, Linear Algebra Appl. 151 (1991) 1–15.
[15] Y. Tian, The Moore–Penrose inverses of $m \times n$ block matrices and their applications, Linear Algebra Appl. 283 (1998) 35–60.
[16] R. Penrose, A generalized inverse for matrices, Proc. Cambridge Philos. Soc. 51 (1955) 406–413.
[17] O. Schabenberger, C. Gotway, Statistical Methods for Spatial Data Analysis, Chapman & Hall/CRC, London, 2005.
[18] A. Gelfand, S. Banerjee, Handbook of Spatial Statistics, Chapman & Hall/CRC, London, 2010, pp. 495–515.
[19] R. Furrer, M. Genton, D. Nychka, Covariance tapering for interpolation of large spatial datasets, J. Comput. Graph. Statist. 15 (2006) 502–523.
[20] R. Furrer, S. Sain, Spatial model fitting for large datasets with applications to climate and microarray problems, Statist. Comput. 19 (2009) 113–128.
[21] N. Cressie, G. Johannesson, Fixed rank kriging for very large spatial data sets, J. Roy. Statist. Soc. Ser. B 70 (2008) 209–226.
[22] S. Banerjee, A. Gelfand, A. Finley, H. Sang, Gaussian predictive process models for large spatial data sets, J. Roy. Statist. Soc. Ser. B 70 (2008) 825–848.
[23] H. Rue, L. Held, Gaussian Markov Random Fields: Theory and Applications, Chapman & Hall/CRC, London, 2005.

Paper III

Sequential Spatial Analysis of Large Datasets with Applications to Modern Earthwork Compaction Roller Measurement Values

Daniel K. Heersink & Reinhard Furrer

Paper submitted to *Spatial Statistics*.

Sequential Spatial Analysis of Large Datasets with Applications to Modern Earthwork Compaction Roller Measurement Values

Daniel K. Heersink[1], Reinhard Furrer*

Institute of Mathematics, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich

Abstract

In the context of road construction, modern earthwork compaction rollers equipped with sensors collect a virtually continuous flow of soil property measurements. This sequential, spatial data can be utilized to improve the quality control of the compaction process through the introduction of intelligent compaction. These roller measurement values are observed indirectly through non-linear measurement operators, non-stationary, inherently multivariate with complex correlation structures, and collected in huge quantities. The problem of modeling and estimation in a spatially correlated setting with large amounts of data is well known and many approaches can be found in the literature. Very few studies have been completed investigating sequential, spatially correlated data outside of a point process framework. We propose a sequential, spatial mixed-effects model and develop a sequential, spatial backfitting algorithm to estimate fixed effects and several independent, spatially correlated processes. This new algorithm is demonstrated in a simulation study and applied to earthwork compaction data.

Keywords: backfitting, sequential modeling, large datasets, hierarchical multivariate spatial models.

2000 MSC: 37M10, 62L12, 62M20, 62M30

*Corresponding author

Email addresses: daniel.heersink@math.uzh.ch (Daniel K. Heersink), reinhard.furrer@math.uzh.ch (Reinhard Furrer)

[1]The contents of this paper are part of the first author's Ph.D. dissertation supervised by the second author.

1. Introduction

Sequential, spatial processes are random process in space and time that occur in a consecutive manner. Such processes are often complex and care is needed in modeling to ensure identifiability of model parameters and maintain computational feasibility. To maintain computational feasibility of a model, researchers must limit the generality of the statistical model to handle the complex nature of such data. Complexities in the modeling stem from data observed indirectly through non-linear measurement operators, non-stationary data, and/or inherently multivariate data with sequential, complex correlation structures. To add to the complexity, the data is usually collected in huge quantities.

An example of such data is modern earthwork compaction as employed for road construction. Modern earthwork compaction rollers are equipped with a sensor measuring soil density and providing data in a virtually continuous flow. These data, coupled with on-board GPS measurements, together termed the roller measurement value (RMV), are large, dense, and spatially correlated. The data are also considered a sequential spatial process as the construction process requires several layers of material be compacted, one on top of another, consecutively.

Many researchers have proposed dynamical spatial models and estimation procedures. Waller et al. [1] develop spatio-temporal models for lung cancer mortality. Wikle et al. [2] examine climatological data in a spatio-temporal framework. Bailey and Krzanowski [3] give an overview of several approaches to multivariate geostatistical data and Christensen and Sain [4] apply latent variable modeling to climate models. Cressie and Wikle [5] propose a very general hierarchical, dynamical spatio-temporal model (DSTM) setup. For more, see the bibliographic notes of Section 7.8 of their book.

Earthwork compaction and RMVs are a novel and challenging application of spatio-temporal modeling. There is a desire to model multiple ranges of spatial variation in RMVs. Due to the unique, sequential nature of modern earthwork compaction, previous spatio-temporal models developed in the literature do not adequately represent the process, and its multiple ranges of spatial variation, under consideration. Data and process models mentioned in Section 7.8 of [5] are too distinct from the current application and a new model must be developed.

A general, sequential, spatial mixed-effects model is proposed. Strategies for fitting

such a model and estimating model parameters for large amounts of data are discussed. There are two aspects to this estimation. First, sparse matrices and finite-range covariances can be utilized to deal with large covariance structures as in [6]. Second, many of the covariance matrices in this model can be written in quasi-Kronecker structure as in [7], greatly decreasing the computation requirements of inverse calculations.

The complex, sequential, spatial data collected from earthwork compaction is presented in Section 2. Section 3 introduces a sequential, spatial mixed-effects model and Section 4 develops two backfitting algorithms for such models. Calculation and computation of covariance structures is also addressed. A simulation study of the backfitting algorithm is presented in Section 5. The algorithm is then applied to earthwork compaction data in Section 6. Finally, in Section 7, current and future research outlook is discussed.

2. Modern earthwork compaction roller measurement values

Modern earthwork compaction rollers are employed to compact material during road construction. A vibrating drum with a diameter of approximately 1m and length of approximately 2m traverses the compaction site in several adjacent driving lanes. These rollers collect compaction and location data as they operate. This data, termed the roller measurement value (RMV), can be modeled as a spatial process. The goal of such a model is to improve the quality of the compaction process. This is achieved by improving the identification of weak, or soft spots, and by ensuring a homogeneous compaction.

The first modern earthwork compaction rollers designed for continuous compaction control (CCC) were used for construction starting in the 1970s in the European community. CCC is a method of documenting compaction and is used to achieve homogeneous compaction in a minimum time [8]. Rudimentary intelligent compaction (IC) technology was first available in the late 1990s. IC is an automated system that adjusts roller operation parameters for optimal compaction based on CCC data [9].

Roller manufacturer Ammann developed their measurement value in the late 1990s. Ammann uses the drum/soil assembly model depicted in Figure 1, where m_f and m_d are the masses of the frame and the drum respectively, g is the acceleration of gravity, z_d and $\ddot{z}_d$ are the drum displacement and acceleration, $m_0 e_0$ is the eccentric mass moment,

Ω is the excitation frequency, and F_s is the drum/soil contact force, to calculate the soil stiffness value (k_s) [10]. Equating the forces of motion of the roller and solving for k_s, ignoring frame inertia, when the drum velocity is zero yields [11]:

$$k_s := \frac{F_s}{z_d} = \Omega^2 \left[m_d + \frac{m_0 e_0 \cos(\Omega t)}{z_d} \right].$$

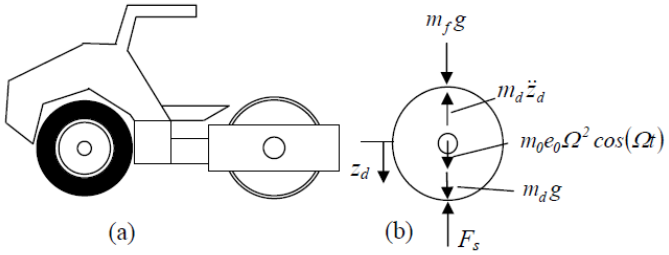

Figure 1: (a) Model of drum/soil assembly for calculation of k_s. (b) Free body diagram of forces involved. Figure courtesy of [12].

This non-linear measurement operator provides an aggregate measurement of the underlying soil stiffness. The point measurement is an aggregate of the approximately 1m^3 bulb of material directly below the roller drum, 2m long, 0.5m wide, and approximately 1m into the soil [13].

2.1. Uncertainty in RMVs

RMVs are calculated from frequency and excitation force amplitude information gathered from sensors within the vibrating and rotating smooth drum. The k_s RMVs discussed in this paper are proportional to the stiffness of the underlying soil [14] and directly represent soil stiffness by calculating the contact force and displacement during vibration [15], [16]. Many studies and models have verified that these RMVs represent soil stiffness [17], [18], [11], [19].

RMVs reflect the aggregate stiffness of the underlying material to approximately 1m [13]. Due to this complication of multiple layers of material, the RMVs are often correlated to industry standard spot tests [20].

There are two values of uncertainty in RMVs. The uncertainty in a single RMV is σ_1 and the uncertainty in the difference between two RMVs collected at the same spatial location is σ_2 [14]. Of importance to the current exposition is σ_1.

2.2. Modeling concerns of RMVs

The inherent uncertainty of RMVs lends itself to a statistical modeling approach. Important attributes of RMVs to consider when modeling include they are indirectly observed through non-linear measurement operators and are inherently multivariate with complex correlation structures. They are also collected in large quantities, creating computational concerns.

The nature of RMV data collection requires a sequential model. There are few sequential spatial models and estimation procedures proposed in the literature and those are limited or do not have the necessary flexibility and generality required to model RMVs. van Lieshout [21] discusses sequential spatial models in an image processing context using Markov marked point processes.

The sequential nature of the compaction process introduces non-stationarity into the data. The roller compacts consecutive layers of material, but records an aggregate measure of stiffness that includes multiple layers of material.

These RMVs can be used to improve IC. They are also vitally important for improvement of quality assurance (QA) and quality control (QC). To implement such improvements, a stochastic model of the RMVs must be developed and estimation procedures for parameters in the model need to be developed and implemented. A model of each layer, a model of the layer-to-layer process, and a dynamical model of the compaction process must be developed. A model of each layer is developed in [22]. This paper extends that model and addresses the layer-to-layer modeling process and estimation.

3. A sequential, spatial mixed-effects model

Let $\{y(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{D} \subset \mathbb{R}^d\}$ be a spatial process. Assume further that the process can be decomposed into a mean process, a spatial process, and an error process: $y(\boldsymbol{s}) = \boldsymbol{x}(\boldsymbol{s})^T\boldsymbol{\beta} + \alpha(\boldsymbol{s}) + \varepsilon(\boldsymbol{s})$. Each term of the decomposition measures a different scale of variation [23]. This decomposition leads to the additive, spatial mixed-effects model described by the n-vector $\boldsymbol{y}$ and can be written as

$$\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\alpha} + \boldsymbol{\varepsilon}, \tag{1}$$

where $\mathbf{X}\boldsymbol{\beta}$ represents the fixed effects and $\mathbf{X}$ is the design matrix of full rank; $\boldsymbol{\alpha}$ represents a random, zero-mean spatial process with $\mathrm{Var}(\boldsymbol{\alpha}) = \boldsymbol{\Sigma}(\boldsymbol{\theta})$, a spatial covariance matrix parameterized by $\boldsymbol{\theta}$; $\boldsymbol{\varepsilon}$ represents a random, zero-mean error process with $\mathrm{Var}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}$, orthogonal to $\boldsymbol{\alpha}$. This model leads to:

$$\mathrm{Var}(\boldsymbol{y}) = \boldsymbol{\Sigma}_{\boldsymbol{y}} = \boldsymbol{\Sigma}(\boldsymbol{\theta}) + \sigma^2\mathbf{I}, \tag{2}$$

a symmetric, positive definite matrix.

If the covariance parameters $\boldsymbol{\theta}$ and σ are known, estimates of the fixed effects $\boldsymbol{\beta}$ and the spatial random effects follow directly as [6]:

$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}^T\boldsymbol{\Sigma}_{\boldsymbol{y}}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\boldsymbol{\Sigma}_{\boldsymbol{y}}^{-1}\boldsymbol{y} \quad \text{(generalized least-squares estimator)}, \tag{3}$$

$$\widehat{\boldsymbol{\alpha}} = \boldsymbol{\Sigma}(\boldsymbol{\theta})\boldsymbol{\Sigma}_{\boldsymbol{y}}^{-1}\left(\boldsymbol{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}\right) \quad \text{(spatial smoothing)}. \tag{4}$$

The covariance parameters are rarely known in practice though and they need to be estimated by e.g. maximum likelihood, method of moments, etc. Once appropriate estimates of these parameters have been made, the estimated covariance matrices may be plugged into (3) and (4).

The first extension to model (1) is to allow for multiple spatial processes. Let $\boldsymbol{\alpha}_1, \ldots \boldsymbol{\alpha}_k$ be k, independent spatial processes, independent of $\boldsymbol{\varepsilon}$, each corresponding to a different scale of variance. The observation vector can then be decomposed as follows:

$$\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\alpha}_1 + \cdots + \boldsymbol{\alpha}_k + \boldsymbol{\varepsilon}. \tag{5}$$

This model leads to a covariance structure of $\mathrm{Var}(\boldsymbol{y}) = \boldsymbol{\Sigma}_{\boldsymbol{y}} = \boldsymbol{\Sigma}_1 + \cdots + \boldsymbol{\Sigma}_k + \boldsymbol{\Sigma}_\varepsilon$. Here, $\mathrm{Var}(\boldsymbol{\alpha}_j) = \boldsymbol{\Sigma}_j(\boldsymbol{\theta}_j) = \boldsymbol{\Sigma}_j$, where we have dropped the explicit $\boldsymbol{\theta}_j$ parameter representation for simplicity. Cressie and Wikle [5] term model (5) the components-of-variation model. As detailed there in Section 6.6, care needs to be taken with such a model to ensure identifiability of all covariance parameters $\boldsymbol{\theta}_j$.

3.1. Sequential model formulation

Complex processes like modern earthwork compaction, where we have sequential observations of multiple spatial processes requires a more complex model than (1) or (5). These models can be extended to a sequential model by generalizing to p different observation vectors $\boldsymbol{y}_t$ as

$$\begin{aligned}
\boldsymbol{y}_1 &= \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\alpha}_1 + \boldsymbol{\varepsilon}_1 \\
\boldsymbol{y}_2 &= \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2 + \boldsymbol{\varepsilon}_2 \\
&\vdots \\
\boldsymbol{y}_p &= \mathbf{X}_p\boldsymbol{\beta}_p + \boldsymbol{\alpha}_1 + \cdots + \boldsymbol{\alpha}_p + \boldsymbol{\varepsilon}_p,
\end{aligned} \tag{6}$$

where the $\boldsymbol{\alpha}_t$s correspond to a new spatial process observed at time t and the $\boldsymbol{\varepsilon}_t$s are mutually independent of the $\boldsymbol{\alpha}_t$s. Thus, $\mathbf{Y} = (\boldsymbol{y}_1^T, \ldots, \boldsymbol{y}_p^T)^T$ has a full block covariance structure: $\mathrm{Var}(\mathbf{Y}) = \boldsymbol{\Sigma}_{\boldsymbol{\alpha}} + \mathrm{bdiag}(\sigma_1^2\mathbf{I}, \ldots, \sigma_p^2\mathbf{I})$, where $\mathrm{bdiag}(\mathbf{A}_1, \ldots, \mathbf{A}_k)$ is a block diagonal matrix with $\mathbf{A}_1, \ldots, \mathbf{A}_k$ as diagonal blocks and

$$\boldsymbol{\Sigma}_{\boldsymbol{\alpha}} = \begin{bmatrix}
\boldsymbol{\Sigma}_1 & \boldsymbol{\Sigma}_1 & \ldots & \boldsymbol{\Sigma}_1 \\
\boldsymbol{\Sigma}_1 & \boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2 & \ldots & \boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2 \\
\vdots & \vdots & \ddots & \vdots \\
\boldsymbol{\Sigma}_1 & \boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2 & \ldots & \sum_{t=1}^{p} \boldsymbol{\Sigma}_t
\end{bmatrix}. \tag{7}$$

More complex correlation structures can be modeled in a straightforward generalization to

$$\mathbf{Y} = \begin{bmatrix} \mathbf{X}_1\boldsymbol{\beta}_1 \\ \vdots \\ \mathbf{X}_p\boldsymbol{\beta}_p \end{bmatrix} + (\mathbf{C} \otimes \mathbf{I}) \begin{bmatrix} \boldsymbol{\alpha}_1 \\ \vdots \\ \boldsymbol{\alpha}_p \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \vdots \\ \boldsymbol{\varepsilon}_p \end{bmatrix},$$

where $\mathbf{C}$ is an arbitrary matrix defining how the $\boldsymbol{\alpha}_t$s are modeled for each $\boldsymbol{y}_t$ and $\otimes$ is the Kronecker product. This can be written more simply as

$$\mathbf{Y} = \mathcal{X}\boldsymbol{\beta} + \mathcal{C}\boldsymbol{\alpha} + \boldsymbol{\varepsilon}, \tag{8}$$

where $\mathcal{X} = \mathrm{bdiag}(\mathbf{X}_1, \ldots, \mathbf{X}_p)$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \ldots, \boldsymbol{\beta}_p^T)^T$, $\mathcal{C} = \mathbf{C} \otimes \mathbf{I}$, $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1^T, \ldots, \boldsymbol{\alpha}_p^T)^T$, and

$\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1^T, \ldots, \boldsymbol{\varepsilon}_p^T)^T$. The associated variance of this model is then

$$\text{Var}(\mathbf{Y}) = \boldsymbol{\Sigma}_{\mathbf{Y}} = \mathcal{C}\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}\mathcal{C}^T + \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}, \tag{9}$$

where $\boldsymbol{\Sigma}_{\boldsymbol{\alpha}} = \text{bdiag}(\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_p)$ and $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}} = \text{bdiag}(\sigma_1^2\mathbf{I}, \ldots, \sigma_p^2\mathbf{I})$.

3.2. State-Space model formulation

The previous developments assume that each observation vector, $\boldsymbol{y}_t$, is observed at the same set of locations. In practice, this may not be the case and each $\boldsymbol{y}_t$ could be observed at a unique set of locations. Reformulating the problem in a state-space formulation addresses this problem.

Assume the process, $\mathbf{Z} = (\boldsymbol{z}_1^T, \ldots, \boldsymbol{z}_p^T)^T$, is defined on a regular grid with the additive decomposition:

$$\mathbf{Z} = \mathcal{X}\boldsymbol{\beta} + \mathcal{C}\boldsymbol{\alpha}. \tag{10}$$

The observations of the process can then be modeled as $\boldsymbol{y}_t = \mathbf{H}_t\boldsymbol{z}_t + \boldsymbol{\varepsilon}_t$. Here, $\mathbf{H}_t$ is the matrix that maps the grid points, to the observed locations. Thus we can write, $\mathbf{Y} = (\boldsymbol{y}_1^T, \ldots, \boldsymbol{y}_p^T)^T = \mathbf{H}\mathbf{Z} + \boldsymbol{\varepsilon}$, where $\mathbf{H} = \text{bdiag}(\mathbf{H}_1, \ldots, \mathbf{H}_p)$, and the variance of the observed vector is:

$$\text{Var}(\mathbf{Y}) = \boldsymbol{\Sigma}_{\mathbf{Y}} = \mathbf{H}\mathcal{C}\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}\mathcal{C}^T\mathbf{H}^T + \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}. \tag{11}$$

3.3. Correlated structures

Under models (6) and (8), the estimation of spatial parameters is more difficult than models (1) and (5) as the individual spatial terms are correlated. The convenient block diagonal structure of the covariance matrix created by independent $\boldsymbol{\alpha}_t$s is replaced by a (potentially) full block matrix. When dealing with such complicated correlation structures, computational complexity and the variance of the estimates must be addressed.

In general, (8) and (10) are too complex to handle directly as the estimates of the modeling terms and their associated variances, i.e. (9) and (11), require the inversion of a (potentially) full block matrix. However, certain general model definitions have covariance structures of the form $\text{bdiag}(\mathbf{A}_k) + \boldsymbol{u}\boldsymbol{v}^* \otimes \mathbf{E}$, where $\mathbf{A}_k, k = 1, \ldots, p$, and $\mathbf{E}$ are covariance matrices and $\boldsymbol{u}, \boldsymbol{v}$ are two p-vectors defining the correlation structure, the so-called quasi-Kronecker structure [7]. This special matrix structure allows the calculation

of block matrix inverses on the scale of the individual blocks $\mathbf{A}_k$, thus both estimation and variance issues can be addressed.

If there are only a small number of observations of the field, more complex model structures can be dealt with. For example, sequential, iterative processes of an "autoregressive" type can be incorporated. Let $p = 3$ and

$$\begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \boldsymbol{y}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1\boldsymbol{\beta}_1 \\ \mathbf{X}_2\boldsymbol{\beta}_2 \\ \mathbf{X}_3\boldsymbol{\beta}_3 \end{bmatrix} + \begin{bmatrix} \boldsymbol{\alpha}_1 \\ c\boldsymbol{\alpha}_1 \\ c^2\boldsymbol{\alpha}_1 \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \boldsymbol{\alpha}_2 \\ c\boldsymbol{\alpha}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \boldsymbol{\alpha}_3 \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \boldsymbol{\varepsilon}_3 \end{bmatrix}. \tag{12}$$

Then, $\mathbf{C} = \begin{bmatrix} 1 & 0 & 0 \\ c & 1 & 0 \\ c^2 & c & 1 \end{bmatrix}$ and $\begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \boldsymbol{y}_3 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{X}_1\boldsymbol{\beta}_1 \\ \mathbf{X}_2\boldsymbol{\beta}_2 \\ \mathbf{X}_3\boldsymbol{\beta}_3 \end{bmatrix}, \boldsymbol{\Sigma}_{\mathbf{Y}} \right)$. Where,

$$\begin{aligned} \boldsymbol{\Sigma}_{\mathbf{Y}} &= \begin{bmatrix} \sigma^2\mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \sigma^2\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sigma^2\mathbf{I} + \boldsymbol{\Sigma}_3 \end{bmatrix} + \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ c\mathbf{I} & \mathbf{I} \\ c^2\mathbf{I} & c\mathbf{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_2 \end{bmatrix} \begin{bmatrix} \mathbf{I} & c\mathbf{I} & c^2\mathbf{I} \\ \mathbf{0} & \mathbf{I} & c\mathbf{I} \end{bmatrix} \\ &= \mathbf{A} + \mathbf{U}\mathbf{B}\mathbf{U}^T. \end{aligned} \tag{13}$$

Using the matrix identity $(\mathbf{A} + \mathbf{U}\mathbf{B}\mathbf{U}^T)^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}(\mathbf{B}^{-1} + \mathbf{U}^T\mathbf{A}^{-1}\mathbf{U})^{-1}\mathbf{U}^T\mathbf{A}^{-1}$ [24],

$$\boldsymbol{\Sigma}_{\mathbf{Y}}^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}\mathbf{V}^{-1}\mathbf{U}^T\mathbf{A}^{-1}, \text{ where}$$

$$\mathbf{V} = \text{bdiag}\left(\boldsymbol{\Sigma}_1^{-1} + \frac{1}{\sigma^2}\mathbf{I}, \boldsymbol{\Sigma}_2^{-1} \right) + \begin{bmatrix} c \\ 1 \end{bmatrix} \begin{bmatrix} c & 1 \end{bmatrix} \otimes \left(c^2(\sigma^2\mathbf{I} + \boldsymbol{\Sigma}_3)^{-1} + \frac{1}{\sigma^2}\mathbf{I} \right).$$

Thus, $\mathbf{V}$ is in quasi-Kronecker form and the inverse is easily calculable using equation (4) from [7]. This "autoregressive" type structure can be extended to any p, incorporating a practical aspect of the modeling to sequential data collection scenarios.

Estimation of modeling terms through the backfitting algorithm and their associated variances requires calculation of a (sometimes very) large matrix inverse. Through utilization of the quasi-Kronecker form that $\mathbf{V}$ takes, the inverse is computationally efficient

and only requires calculation of inverses on the order of individual blocks of the entire matrix.

All of these models have covariance structures that require a great deal of care when estimating parameters. The backfitting algorithm is a simple, recursive method well suited to this purpose. The backfitting algorithm is a recursive method, equivalent to the Gauss-Seidel algorithm when the updates to additive components are done sequentially rather than simultaneously, used to obtain solutions to the normal equations in a mixed model framework [25]. The following will be a description of a backfitting algorithm for estimating the modeling terms.

4. A backfitting algorithm for sequential data

The backfitting algorithm has been employed on a wide range of additive models, e.g. [6], [25], [26]. Applied to model (5), the backfitting algorithm then consists of iteratively estimating the fixed effects and the spatial terms as represented below.

Multiple Additive Spatial Terms Backfitting Algorithm

[0] Let $\widehat{\boldsymbol{\alpha}}_i^{(0)}$, for $i = 1, \ldots, k$, be an initial guess and put $j = 0$

[1] $j = j + 1$

[2] $\widehat{\boldsymbol{\beta}}^{(j)} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\left(\boldsymbol{y} - \sum_{i=1}^{k}\widehat{\boldsymbol{\alpha}}_i^{(j-1)}\right)$

[3] For $i = 1, \ldots, k$, estimate covariance parameters to get $\widehat{\sigma}^{2,(j)}$ and $\widehat{\boldsymbol{\Sigma}}_i^{(j)}$, then put $\widehat{\boldsymbol{\alpha}}_i^{(j)} = \widehat{\mathbf{W}}_i^{(j)}\left(\boldsymbol{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}^{(j)} - \sum_{\ell=1}^{i-1}\widehat{\boldsymbol{\alpha}}_\ell^{(j)} - \sum_{\ell=i+1}^{k}\widehat{\boldsymbol{\alpha}}_\ell^{(j-1)}\right)$, $i = 1, \ldots, n$

[4] Repeat [2] to [4] until convergence.

In step [3] of the algorithm, $\widehat{\mathbf{W}}_i^{(j)} = \widehat{\boldsymbol{\Sigma}}_i^{(j)}\left(\widehat{\sigma}^{2,(j)}\mathbf{I} + \widehat{\boldsymbol{\Sigma}}_i^{(j)}\right)^{-1}$. For convenience, we write $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{y},i}^{(j)} = \widehat{\sigma}^{2,(j)}\mathbf{I} + \widehat{\boldsymbol{\Sigma}}_i^{(j)}$.

4.1. Backfitting convergence

Buja et al. [25] proved that the backfitting algorithm is guaranteed to converge. Proof of the validity of the algorithm then requires proof that the iterative ordinary least-squares estimates of $\boldsymbol{\beta}$ are equivalent to the "global" generalized least-squares estimate.

Proof of equivalence utilizes the matrix identity $\mathbf{I}-\mathbf{A}(\mathbf{A}+\lambda\mathbf{I})^{-1} = \lambda(\mathbf{A}+\lambda\mathbf{I})^{-1}$, where $\mathbf{A}$ is semi-positive definite and $\lambda > 0$.

First, let $k = 1$. Then, $\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\alpha}_1 + \boldsymbol{\varepsilon}$, where $\boldsymbol{\alpha}_1$ is a zero-mean spatial process. This is the model investigated in [6] and the proof is found therein.

Next, let $k = 2$. Then, $\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2 + \boldsymbol{\varepsilon}$. After convergence of the backfitting procedure we write $\widehat{\boldsymbol{\alpha}}_1 = \widehat{\boldsymbol{\Sigma}}_1\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{y},1}^{-1}\left(\boldsymbol{y} - \mathbf{X}\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\alpha}}_2\right)$ and $\widehat{\boldsymbol{\alpha}}_2 = \widehat{\boldsymbol{\Sigma}}_2\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{y},2}^{-1}\left(\boldsymbol{y} - \mathbf{X}\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\alpha}}_1\right)$. Then,

$$\begin{aligned}
\mathbf{X}^T\mathbf{X}\widehat{\boldsymbol{\beta}} &= \mathbf{X}^T\left(\boldsymbol{y} - \widehat{\boldsymbol{\alpha}}_1 - \widehat{\boldsymbol{\alpha}}_2\right), \\
\widehat{\boldsymbol{\alpha}}_1 &= \widehat{\boldsymbol{\Sigma}}_1\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{y},1}^{-1}\left(\boldsymbol{y} - \mathbf{X}\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\alpha}}_2\right), \text{ and} \\
\widehat{\boldsymbol{\alpha}}_2 &= \widehat{\boldsymbol{\Sigma}}_2\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{y},2}^{-1}\left(\boldsymbol{y} - \mathbf{X}\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\alpha}}_1\right)
\end{aligned}$$

Solving for $\widehat{\boldsymbol{\alpha}}_1$ and $\widehat{\boldsymbol{\alpha}}_2$ in terms of $\widehat{\boldsymbol{\beta}}$ yields:

$$\begin{aligned}
\widehat{\boldsymbol{\alpha}}_1 &= \mathbf{Q}\left(\boldsymbol{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}\right) \text{ and} \\
\widehat{\boldsymbol{\alpha}}_2 &= \widehat{\boldsymbol{\Sigma}}_2\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{y},2}^{-1}(\mathbf{I} - \mathbf{Q})\left(\boldsymbol{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}\right),
\end{aligned}$$

where $\mathbf{Q} = \left(\mathbf{I} - \widehat{\boldsymbol{\Sigma}}_1\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{y},1}^{-1}\widehat{\boldsymbol{\Sigma}}_2\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{y},2}^{-1}\right)^{-1}\widehat{\boldsymbol{\Sigma}}_1\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{y},1}^{-1}\left(\mathbf{I} - \widehat{\boldsymbol{\Sigma}}_2\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{y},2}^{-1}\right)$. Thus,

$$\begin{aligned}
\mathbf{X}^T\mathbf{X}\widehat{\boldsymbol{\beta}} &= \mathbf{X}^T\left(\boldsymbol{y} - \mathbf{Q}(\boldsymbol{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}) - \widehat{\boldsymbol{\Sigma}}_2\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{y},2}^{-1}(\mathbf{I} - \mathbf{Q})(\boldsymbol{y} - \mathbf{X}\widehat{\boldsymbol{\beta}})\right) \\
\mathbf{X}^T\left(\mathbf{I} - \mathbf{Q} - \widehat{\boldsymbol{\Sigma}}_2\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{y},2}^{-1}(\mathbf{I} - \mathbf{Q})\right)\mathbf{X}\widehat{\boldsymbol{\beta}} &= \mathbf{X}^T\left(\mathbf{I} - \mathbf{Q} - \widehat{\boldsymbol{\Sigma}}_2\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{y},2}^{-1}(\mathbf{I} - \mathbf{Q})\right)\boldsymbol{y} \\
\mathbf{X}^T\left(\left(\mathbf{I} - \widehat{\boldsymbol{\Sigma}}_2\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{y},2}^{-1}\right)(\mathbf{I} - \mathbf{Q})\right)\mathbf{X}\widehat{\boldsymbol{\beta}} &= \mathbf{X}^T\left(\left(\mathbf{I} - \widehat{\boldsymbol{\Sigma}}_2\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{y},2}^{-1}\right)(\mathbf{I} - \mathbf{Q})\right)\boldsymbol{y} \\
\mathbf{X}^T\left(\widehat{\sigma}^2\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{y},2}^{-1}(\mathbf{I} - \mathbf{Q})\right)\mathbf{X}\widehat{\boldsymbol{\beta}} &= \mathbf{X}^T\left(\widehat{\sigma}^2\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{y},2}^{-1}(\mathbf{I} - \mathbf{Q})\right)\boldsymbol{y} \\
\mathbf{X}^T\boldsymbol{\Omega}^{-1}\mathbf{X}\widehat{\boldsymbol{\beta}} &= \mathbf{X}^T\boldsymbol{\Omega}^{-1}\boldsymbol{y}.
\end{aligned}$$

Now, $\mathbf{I} - \mathbf{Q} = \left(\mathbf{I} - \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{y},1}\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{y},2}^{-1} - \widehat{\sigma}^2\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{y},2}^{-1}\right)^{-1}$. Therefore, $\boldsymbol{\Omega}^{-1} = \left(\widehat{\boldsymbol{\Sigma}}_1 + \widehat{\boldsymbol{\Sigma}}_2 + \widehat{\sigma}^2\mathbf{I}\right)^{-1}$, and $\widehat{\boldsymbol{\beta}}$

is the estimated generalized least-squares estimate.

For values of $k > 2$, the proof of equivalence is similar and therefore omitted here.

4.2. Sequential observations

When dealing with multiple observations in a sequential setting like that of model (6), the backfitting algorithm must be modified. The modified backfitting algorithm consists of p sequential steps of iteratively estimating the fixed effects, $\boldsymbol{\beta}_t$, using ordinary least squares and the spatial terms, $\boldsymbol{\alpha}_t, t = 1, \ldots, p$ using universal kriging as represented below.

Sequential Backfitting Algorithm

[$0^\dagger$] For $t = 1$ to p

[$1^\dagger$] Let $\widehat{\boldsymbol{\alpha}}_t^{(0)}$ be an initial guess and put $j = 0$

[$2^\dagger$] $j = j + 1$

[$3^\dagger$] $\widehat{\boldsymbol{\beta}}_t^{(j)} = \left(\mathbf{X}_t^T \mathbf{X}_t\right)^{-1} \mathbf{X}_t^T \left(\boldsymbol{y}_t - \sum_{i=1}^{t} \widehat{\boldsymbol{\alpha}}_i^{(j-1)}\right)$

[$4^\dagger$] Estimate covariance parameters to get $\widehat{\sigma}_t^{2,(j)}$ and $\widehat{\boldsymbol{\Sigma}}_t^{(j)}$, then put
$\widehat{\boldsymbol{\alpha}}_t^{(j)} = \widehat{\mathbf{W}}_t^{(j)} \left(\boldsymbol{y}_t - \mathbf{X}_t \widehat{\boldsymbol{\beta}}_t^{(j)} - \sum_{\ell=1}^{t-1} \widehat{\boldsymbol{\alpha}}_\ell^{(j)}\right)$

[$5^\dagger$] Repeat [$3^\dagger$] to [$5^\dagger$] until convergence.

In step [$4^\dagger$] of the algorithm,

$$\widehat{\mathbf{W}}_t^{(j)} = \widehat{\boldsymbol{\Sigma}}_t^{(j)} \left(\widehat{\sigma}_t^{2,(j)} \mathbf{I} + \widehat{\boldsymbol{\Sigma}}_t^{(j)}\right)^{-1} = \widehat{\boldsymbol{\Sigma}}_t^{(j)} \left(\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{y},t}^{(j)}\right)^{-1}.$$

Each sequential step t of the algorithm is equivalent to the Multiple Additive Spatial Terms Backfitting Algorithm.

5. Simulation study

A simulation study was performed to numerically demonstrate convergence of the Sequential Backfitting Algorithm. Six sequential spatial fields were simulated on a uniform

grid as an extension of model (12). The algorithm was then applied to the simulated data. The uniform grid chosen has 81 x-direction values and 40 y-direction values on the interval $[-2, 2]$. The greater density in the x-direction is to simulate the physical process of roller compaction. The fixed effects were taken to be $\boldsymbol{\beta} = (10, 2.5, -3.5)^T$, where β_0 corresponds to a constant, β_1 is the x-coordinate, and β_2 is a covariate defining direction of movement (0 for left-to-right and 1 for right-to-left). The constant c term was taken to be 0.5 for the simulation. The backfitting algorithm was then run using different values of c to investigate the importance of this parameter.

All six spatial fields were chosen to have a spherical covariance structure (i.e. for a given lag h, the covariance function is defined as:

$$C(h;\boldsymbol{\theta}) = \begin{cases} \theta_0 + \theta_2 & \text{for} \quad h = 0 \\ \theta_0(1 - 1.5(h/\theta_1) + .5(h/\theta_1)^3) & \text{for} \quad h \in (0, \theta_1) \\ 0 & \text{for} \quad h \geq \theta_1 \end{cases}, \tag{14}$$

where θ_0 is the partial sill, θ_1 is the range, and θ_2 is the nugget of the spatial process and corresponds to the variance of the $\boldsymbol{\varepsilon}_t$s.

For the simulation, $\boldsymbol{\theta}_1 = (4.0, 0.5, 1.0)^T$, $\boldsymbol{\theta}_2 = (8.0, 0.5, 1.0)^T$, $\boldsymbol{\theta}_3 = (12, 0.5, 1.0)^T$, $\boldsymbol{\theta}_4 = (12, 0.5, 0.75)$, $\boldsymbol{\theta}_5 = (12, 0.5, 0.5)$, and $\boldsymbol{\theta}_6 = (12, 0.3, 0.5)$. This sequence represents an increasing partial sill, decreasing the relative amount of information available in successive layers; then a decreasing nugget, decreasing the noise in successive layers; and finally a decreasing range, again decreasing the relative amount of information available.

Following the work of Genton [27], a weighted least squares procedure utilizing the weights proposed by Cressie [23] was used to estimate the covariance parameters from a robust empirical variogram of the residuals. The algorithm estimates the fixed effects almost exactly and the mean estimated fixed effects and corresponding standard deviations of these estimates are the same for all values of c: $\widehat{\boldsymbol{\beta}} = (10.01, 2.51, -3.5)^T$, with a simulation standard deviation of $sd(\widehat{\boldsymbol{\beta}}) = (0.26, 0.21, 0.09)^T$. The estimated spatial parameters from the simulation study are presented in Table 1. The first row of the table reproduces the true values. The estimation of $\boldsymbol{\theta}_1$ is independent of the value of c, since the parameter is absent from the estimating equations. $\widehat{\boldsymbol{\theta}}_1 = (3.78, 0.51, 1.12)^T$, and is omitted from the table, $sd(\widehat{\boldsymbol{\theta}}_1) = (0.3, 0.01, 0.1)^T$.

Table 1: Results of 100 simulations of an extension to six layers of model (12) using the Sequential Backfitting Algorithm. Results reported in the Table are mean values of the 100 simulations.

c	$\widehat{\boldsymbol{\theta}}_2^T$	$\widehat{\boldsymbol{\theta}}_3^T$	$\widehat{\boldsymbol{\theta}}_4^T$	$\widehat{\boldsymbol{\theta}}_5^T$	$\widehat{\boldsymbol{\theta}}_6^T$
	$(8.0, 0.5, 1.0)$	$(12, 0.5, 1.0)$	$(12, 0.5, 0.75)$	$(12, 0.5, 0.5)$	$(12, 0.3, 0.5)$
0.0	$(8.72, 0.52, 1.35)$	$(14.09, 0.54, 1.64)$	$(15.13, 0.52, 1.45)$	$(15.16, 0.53, 1.30)$	$(14.13, 0.37, 2.18)$
0.1	$(8.40, 0.52, 1.36)$	$(13.38, 0.53, 1.68)$	$(13.89, 0.52, 1.45)$	$(13.81, 0.53, 1.29)$	$(12.91, 0.35, 2.11)$
0.2	$(8.15, 0.52, 1.37)$	$(12.81, 0.53, 1.69)$	$(12.93, 0.52, 1.45)$	$(12.75, 0.53, 1.28)$	$(11.98, 0.34, 2.05)$
0.3	$(7.98, 0.52, 1.37)$	$(12.36, 0.53, 1.71)$	$(12.24, 0.52, 1.45)$	$(11.99, 0.53, 1.27)$	$(11.32, 0.33, 2.00)$
0.4	$(7.87, 0.52, 1.37)$	$(12.04, 0.53, 1.72)$	$(11.81, 0.52, 1.44)$	$(11.51, 0.53, 1.25)$	$(10.93, 0.33, 1.96)$
0.5	$(7.83, 0.52, 1.37)$	$(11.86, 0.53, 1.72)$	$(11.65, 0.52, 1.43)$	$(11.34, 0.53, 1.23)$	$(10.80, 0.33, 1.93)$
0.6	$(7.86, 0.52, 1.36)$	$(11.82, 0.53, 1.71)$	$(11.76, 0.52, 1.41)$	$(11.45, 0.53, 1.20)$	$(10.92, 0.33, 1.91)$
0.7	$(7.96, 0.52, 1.36)$	$(11.93, 0.53, 1.70)$	$(12.14, 0.52, 1.38)$	$(11.87, 0.53, 1.19)$	$(11.32, 0.34, 1.91)$
0.8	$(8.13, 0.52, 1.35)$	$(12.18, 0.53, 1.68)$	$(12.80, 0.53, 1.36)$	$(12.59, 0.53, 1.18)$	$(12.00, 0.34, 1.94)$
0.9	$(8.37, 0.52, 1.34)$	$(12.59, 0.53, 1.66)$	$(13.74, 0.53, 1.34)$	$(13.61, 0.53, 1.20)$	$(12.97, 0.36, 2.01)$
1.0	$(8.68, 0.52, 1.33)$	$(13.17, 0.54, 1.64)$	$(14.96, 0.53, 1.34)$	$(14.95, 0.53, 1.25)$	$(14.25, 0.37, 2.13)$

Figure 2 depicts plots of the relative bias of the range and the total sill for the spatial effects parameter estimates $\widehat{\boldsymbol{\theta}}_1, \ldots, \widehat{\boldsymbol{\theta}}_6$, with associated standard deviations. As can be seen in the figures, the relative bias of the range is essentially constant for $\theta_1 = 0.5$ and is minimized at the true value of c for smaller values. The relative bias of the total sill is minimized at the true value of c. The "lower standard deviation bound" of all layers is greater than zero at $c = 0$, thus necessitating the inclusion of the parameter in the algorithm. The lower bound is also greater than zero at $c = 1$, thus some information about the true value of the parameter must be obtained to minimize the bias of the estimates.

As can be seen in Table 1 and Figure 2, estimation of spatial parameters is relatively unbiased, with small variance. Estimation of the fixed effects parameters is also unbiased with small variance. Results of the simulation study thus demonstrate the ability of the algorithm to accurately estimate fixed effects and spatial parameters. The study also demonstrates the true value of the c parameter need not be known, but an approximate value is acceptable for estimation concerns as the relative bias curves have a generally shallow slope.

As θ_0 increases, with θ_1, θ_2 held constant, there is a relative decrease in the amount of information available in a layer. This is evidenced by the increase in the relative bias of successive layers with larger total sills. A decrease in θ_2 while holding θ_0, θ_1 constant, leads to less noisy data, resulting in a decrease in the relative bias. Decreasing θ_1, with θ_0, θ_2 held constant, results in a relative decrease in the amount of information available in a layer. This is evidenced by an increase in the relative bias. The importance of the c parameter can also be seen in the shape of the curves of the total sill. Each successive layer produces a sharper curve, implying greater importance of the accuracy of the c parameter, the more layers observed.

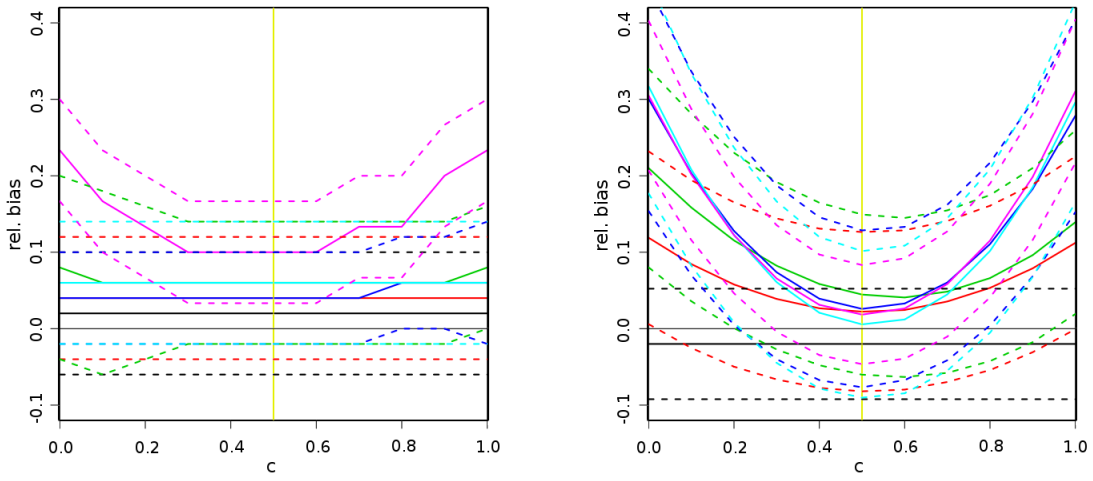

Figure 2: Plot of relative bias of estimated range from the true value (left) and of the relative bias of estimated total sill (right). This relative bias is minimized at the true value of c. Black corresponds to the first layer, red to the second layer, green to the third layer, blue to the fourth layer, cyan to the fifth layer and magenta to the sixth layer. Solid lines are the mean values and the dashed lines correspond to $\pm$ one standard deviation. The yellow vertical line is the true value of the parameter. Black vertical lines at 0 and 1 and horizontal line at 0 are for reference.

6. Application to compaction data

In this section we demonstrate the proposed methodology with a dataset consisting of three layers of RMVs from a test bed at a site in Albertville, MN, USA. The data was collected using a smooth, vibrating drum roller manufactured by Ammann. The test bed lies along a stretch of road adjacent to Interstate 94 ($45°15'45'', -93°42'37''$). The test

bed is approximately 300 meters in length by 15 meters in width and divided into two cells. Measurements from the roller of the subsurface, subgrade, and base layers of the new road under construction were recorded. Figure 3 is a plot of the data from both cells and all three layers.

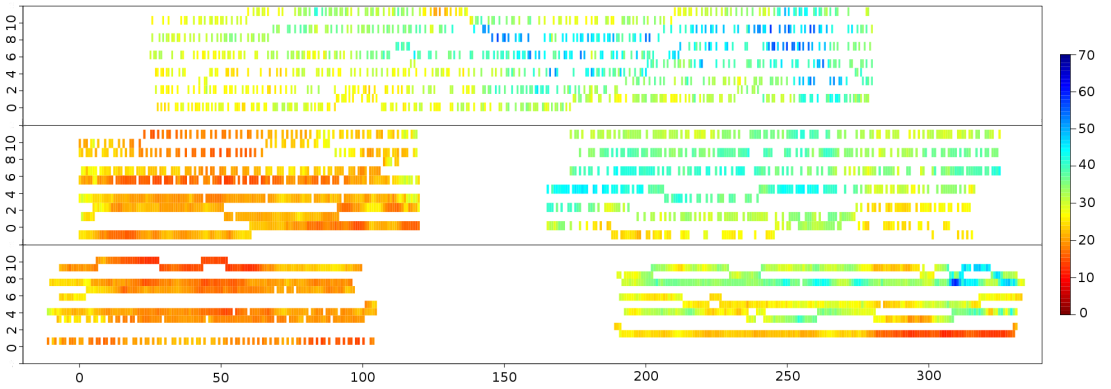

Figure 3: Plot of subsurface (top), subgrade (middle), and base (bottom) data from cells 27 and 28 from the test bed in Albertville, MN, USA. The x-direction is the direction of driving.

Data from each layer was collected separately for each of the two cells, except for the subsurface layer where the roller drove continuously over both cells. Analysis was performed on each cell separately as the standard construction procedure focuses on one cell at a time. The subsurface layer was thus split into two cells at the cell boundary.

The subsurface layer is the existing material at the construction site, whereas the subgrade and base layers are layers of material added during the construction process. The roller traversed the construction site in five to seven lanes, with the coordinates calculated by on-board GPS. Thus, the locations of the observation vector for each layer are unique. This construction process and data collection lends itself nicely to being modeled using equation (12), coupled with the state-space structure of equation (10), where the fixed effects matrix, $\mathbf{X}$, contains an intercept, centered and scaled (x, y)-coordinates of the roller measurement, centered and scaled x-coordinates of the second and third power, and the driving direction of the roller (1 for right-to-left, 0 for left-to-right). Only the first power of the y-coordinates is utilized to minimize the effects of the vast range difference in the x- and y-direction. Polynomials in the y-direction capture a much finer scale of variation than do polynomials in the x-direction.

The process level grid of locations was taken to be a regular grid enclosing all of the

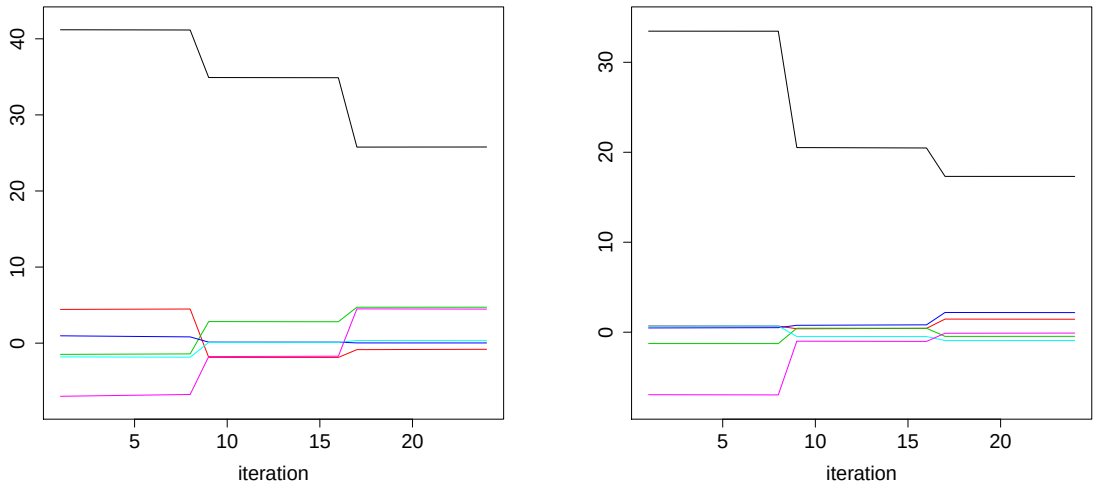

Figure 4: Convergence, measured with the 1-norm, of the fixed effects parameters for cell 27 (left) and cell 28 (right). Cell 27 has a larger mean (black) value than cell 28, which can also be seen in Figure 3. Red corresponds to the x-coordinate, green to the y-coordinate, blue to x^2, cyan to x^3 and magenta to dir. Also see Figure 7.

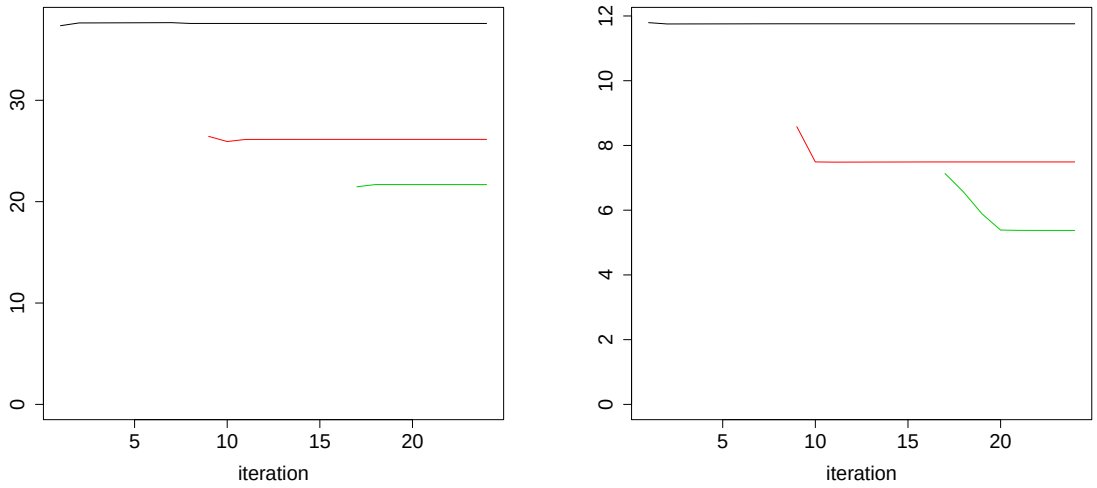

Figure 5: Convergence of the partial sill parameters for cell 27 (left) and cell 28 (right). Cell 28 has much less, approximately half, variation in all three layers than cell 27.

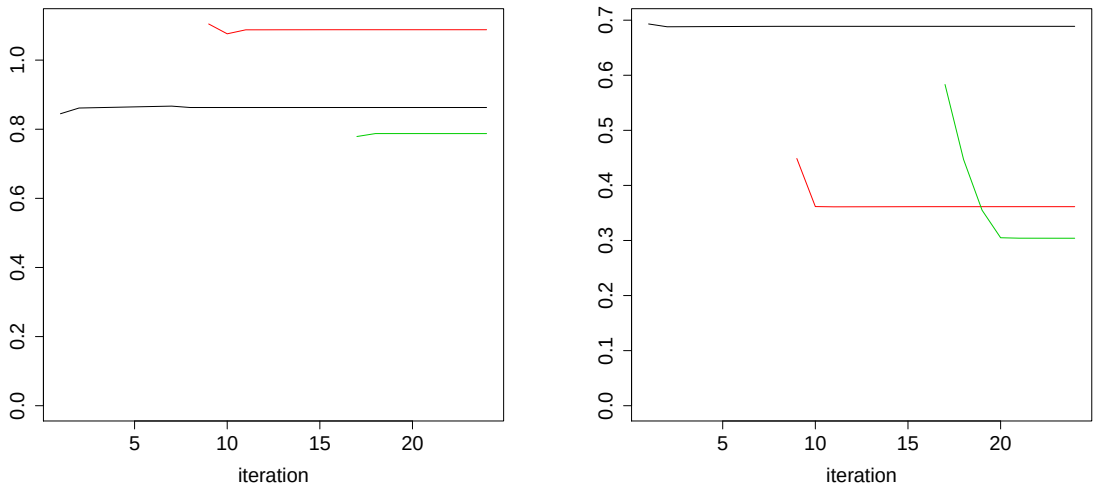

Figure 6: Convergence of the range parameters for cell 27 (left) and cell 28 (right). The spatial range of most of the layers in both cells, excluding the subgrade layer of cell 27, is in the interval (0.3, 0.6), corresponding to a distance of approximately 20-50m. The range of the subgrade layer of cell 27 corresponds to a distance of approximately 75m.

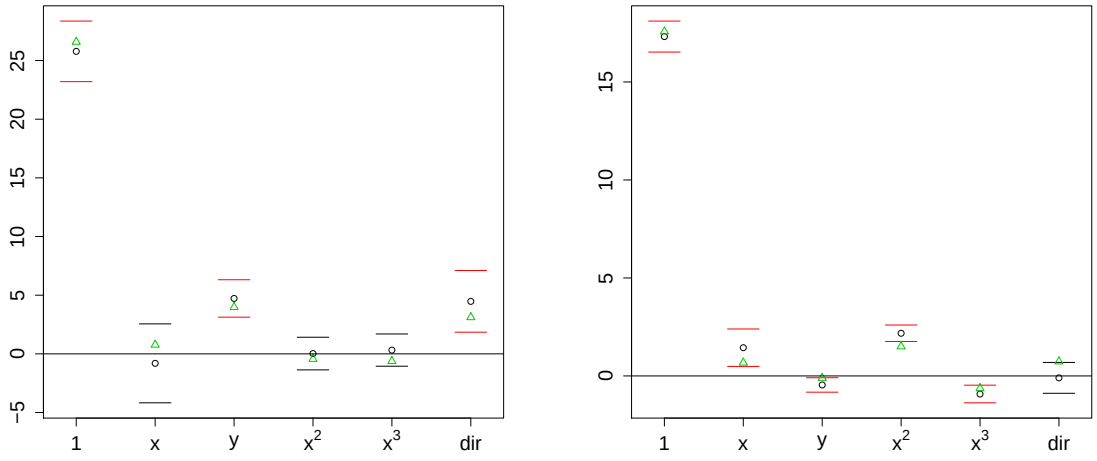

Figure 7: Estimated effects of $\widehat{\boldsymbol{\beta}}$ for cell 27 (left) and cell 28 (right). Backfitting estimates are circles, red signifies significant coefficients. The green triangles are profiled maximum likelihood estimates. The dashes are approximate 95% confidence intervals. x and y are the centered and scaled coordinates and *dir* is the driving direction of the roller.

(centered and scaled) observed locations with forty lanes and driving direction spacing of 0.05. Preliminary variogram estimation of the data supported a covariance structure with finite range. A spherical covariance function was chosen for each layer to generate sparse matrix structures and to decrease computation time.

First, outliers in the data were removed. Outliers were identified as any observation in a driving lane with an absolute difference from a neighboring observation greater than ten. This resulted in thirteen total measurements being removed from the subgrade cells and five total observations being removed from the base cells. The subsurface layer thus contains 517 measurements in cell 27 and 499 measurements in cell 28. The subgrade layer contains 695 measurements in cell 27 and 1613 measurements in cell 28. There are 2882 measurements in cell 27 and 1694 measurements in cell 28 of the base layer.

The Sequential Backfitting Algorithm can then be used to estimate the fixed effects and the covariance parameters of the spatial fields. Here $p = 3$ and initial covariance parameter estimates are chosen using ordinary least squares. Final covariance parameter estimates for cell 27 are $\widehat{\boldsymbol{\theta}}_1 = (37.58, 0.86, 6.55)^T$, $\widehat{\boldsymbol{\theta}}_2 = (26.14, 1.09, 0.57)^T$, $\widehat{\boldsymbol{\theta}}_3 = (21.68, 0.78, 3.37)^T$. Final covariance parameter estimates for cell 28 are $\widehat{\boldsymbol{\theta}}_1 = (11.76, 0.68, 1.52)^T$, $\widehat{\boldsymbol{\theta}}_2 = (7.49, 0.36, 0.78)^T$, $\widehat{\boldsymbol{\theta}}_3 = (5.37, 0.30, 0.18)^T$. Figure 4 depicts the convergence of the fixed effects estimates. The overall stiffness of cell 27 is greater than that of cell 28, which can be seen in the β_0 value. Figures 5 and 6 depict the convergence of the estimates of the partial sill and range parameters for $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3$, respectively. Figure 7 depicts the final estimates of the fixed effects for both cells with approximate 95% confidence intervals using $\text{Var}(\widehat{\boldsymbol{\beta}}) = ((\mathbf{H}_3\mathbf{X}_3)^T(\sigma_3^2\mathbf{I} + \mathbf{H}_3\boldsymbol{\Sigma}_3\mathbf{H}_3^T)^{-1}\mathbf{H}_3\mathbf{X}_3)^{-1}$.

The partial sill parameter estimates are highly dependent on the partial sill estimate of the subsurface. This is an expected result as the roller measures to a depth that "sees" the subsurface in subsequent layers. Excepting the subgrade layer of cell 27, all of the range estimates correspond to a distance of 20–45m. For the subgrade layer of cell 27, the range estimate corresponds to a distance of approximately 75m. With the exception of the subsurface layer of cell 27, all nugget parameter estimates are in the interval (0.1, 1.2). This may indicate an anomalous structure in the subsurface of cell 27. This would also explain the partial sill differences between cell 27 and cell 28.

Maximum likelihood estimation could also be used to estimate the parameters of

this model. The likelihood function is very flat with many local extrema. This leads to numerical difficulties in convergence of the estimates using standard search methods. Results of the simulation study in Section 5 suggest the exact value of the c parameter need not be known. Therefore, c was assumed known and set to 0.5. This is a conservative estimate based on the physics of the process being studied. Each layer of new material is approximately 10cm thick and the roller measures to a depth of approximately 1m. Each successive layer is also a stiffer material than the previous layer, thus most of the measurement is a product of the deepest layers [12].

There is evidence from preliminary variogram estimation that the RMVs are anisotropic [28]. The scaling of the x- and y-coordinates performed in this study eliminated this anisotropy. This is acknowledged to be an ad hoc solution to anisotropy concerns, but the authors feel a suitable coordinate transformation can be performed pre-analysis to limit the influence of this anisotropy. Standard diagnostics show that the residuals are approximately normally distributed and uncorrelated.

7. Discussion and outlook

A very general, sequential, spatial mixed-effects model has been proposed and discussed. This model allows for complex descriptions of the data and correlation structures to be modeled. The backfitting algorithm has been demonstrated to provide computationally efficient estimates of the modeling terms through a simulation study and application to modern earthwork compaction data. Using the quasi-Kronecker matrix form [7], the variance of these estimates is computationally feasible for several complex models. The model proposed is quite general in that an arbitrary number of spatial modeling terms may be added to the model.

This very general model setup allows for a wide range of model descriptions of the data. The zero-mean spatial processes are estimated using the universal kriging estimate (4). Assume instead of a spatial process, a thin-plate-spline (TPS) smoothing approach is desired.

A thin-plate-spline (TPS) is a nonparametric smoothing approach. Consider the problem of estimating a smooth function $f(\boldsymbol{s})$ from n observations $(y_i, \boldsymbol{s}_i)$. Estimation of the

TPS is the problem of finding the function $\widehat{f}$ that minimizes

$$\|\boldsymbol{y} - \mathbf{f}\|^2 + \lambda J_{md}(f),$$

where $\mathbf{f} = (f(\boldsymbol{s}_1), \ldots, f(\boldsymbol{s}_n))^T$ is a smooth function, J_{md} is a penalty functional measuring the "wiggliness" of f and λ is the smoothing parameter controlling the balance between data fitting and smoothness.

This problem is equivalent to minimizing

$$\|\boldsymbol{y} - \mathbf{A}\boldsymbol{\alpha} \ - \mathbf{E}\boldsymbol{\delta}\|^2 + \lambda \boldsymbol{\alpha}^T \mathbf{A}\boldsymbol{\alpha}, \tag{15}$$

with respect to $\boldsymbol{\alpha}, \boldsymbol{\delta}$, subject to $\mathbf{E}^T\boldsymbol{\alpha} = \mathbf{0}$ [29]. Here, $\mathbf{E}$ is the matrix containing the spline basis functions in the null space of J_{md}. For example, if $m = d = 2$ the basis functions are $\phi_1(\boldsymbol{s}) = 1, \phi_2(\boldsymbol{s}) = s_1, \phi_3(\boldsymbol{s}) = s_2$. $\mathbf{A}$ is the matrix containing the remaining basis functions

$$\phi_{md}(r) = \begin{cases} \frac{(-1)^{m+1+d/2}}{2^{2m-1}\pi^{d/2}(m-1)!(m-d/2)!} r^{2m-d}\log(r) & d \text{ even} \\ \\ \frac{\Gamma(d/2-m)}{2^{2m}\pi^{d/2}(m-1)!} r^{2m-d} & d \text{ odd.} \end{cases}$$

Let $\widehat{\boldsymbol{\alpha}}$ be the vector minimizing (15). Then,

$$\begin{aligned} \widehat{\boldsymbol{\alpha}} &= \arg\min_{\boldsymbol{\alpha}} \|\boldsymbol{y} - \mathbf{A}\boldsymbol{\alpha} \ - \mathbf{E}\boldsymbol{\delta}\|^2 + \lambda \boldsymbol{\alpha}^T \mathbf{A}\boldsymbol{\alpha} \\ &= (\mathbf{I} + \lambda\mathbf{A})^{-1}(\boldsymbol{y} - \mathbf{E}\boldsymbol{\delta}) \\ &= \mathbf{A}^{-1}(\lambda\mathbf{I} + \mathbf{A}^{-1})^{-1}(\boldsymbol{y} - \mathbf{E}\boldsymbol{\delta}). \end{aligned}$$

The estimate of a TPS is thus mathematically equivalent to the spatial smoothing of (4). Therefore, TPS terms can be included in the sequential, spatial mixed-effects model instead of random fields and the backfitting algorithms presented will still converge to their proper estimates.

It should be noted that a ridge regression term is also mathematically equivalent and can replace random field components of the model as well.

Thus, the proposed model encompasses a wider range of models where the additive

terms are any spatial process that can be estimated in a mathematically equivalent way to universal kriging.

The correlation structure (13) is only one possible model definition that induces the quasi-Kronecker structure [7]. Here are two examples.

First, let each observation $\boldsymbol{y}_t$, be modeled as the sum of fixed effects, $\mathbf{X}_t\boldsymbol{\beta}$, and two independent spatial processes, $\boldsymbol{\alpha}_t$ and $c_t\boldsymbol{\alpha}$, where c_t is a scaling variable for observation t, and an independent noise process $\boldsymbol{\varepsilon}_t$. Then $\boldsymbol{y}_t = \mathbf{X}_t\boldsymbol{\beta} + \boldsymbol{\alpha}_t + c_t\boldsymbol{\alpha} + \boldsymbol{\varepsilon}_t$. Thus, from (8), $\mathbf{C} = (\mathbf{I}_k, \boldsymbol{c})$, where $\boldsymbol{c} = (c_1, \ldots, c_k)^T$ and $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_k, \boldsymbol{\alpha})^T$. Then, $\boldsymbol{\Sigma}_{\boldsymbol{\alpha}} = \text{bdiag}(\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_k, \boldsymbol{\Sigma}_{\boldsymbol{\alpha}})$. Also, $\mathcal{C}\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}\mathcal{C}^T = \text{bdiag}(\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_k) + \boldsymbol{c}\boldsymbol{c}^T \otimes \boldsymbol{\Sigma}_{\boldsymbol{\alpha}}$. Thus,

$$\text{Var}(\mathbf{Y}) = \text{bdiag}(\boldsymbol{\Sigma}_1 + \sigma^2\mathbf{I}, \ldots, \boldsymbol{\Sigma}_k + \sigma^2\mathbf{I}) + \boldsymbol{c}\boldsymbol{c}^T \otimes \boldsymbol{\Sigma}_{\boldsymbol{\alpha}}, \tag{16}$$

which is in quasi-Kronecker form and the inverse is easily calculable using equation (4) from [7].

Second, let each observation $\boldsymbol{y}_t$, be modeled as the sum of fixed effects, $\mathbf{X}_t\boldsymbol{\beta}$, and scaled, independent spatial processes, $\sum_{j=1}^{k} c_j\boldsymbol{\alpha}_j$, and an independent noise process $\boldsymbol{\varepsilon}_t$. Then $\boldsymbol{y}_t = \mathbf{X}_t\boldsymbol{\beta} + \sum_{j=1}^{n} c_j\boldsymbol{\alpha}_j + \boldsymbol{\varepsilon}_t$. Thus,

$$\mathbf{C} = (c_1, \ldots, c_k) \otimes \mathbf{1} \text{ and } \boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_k)^T.$$

Then, $\boldsymbol{\Sigma}_{\boldsymbol{\alpha}} = \text{bdiag}(\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_k)$ and $\mathcal{C}\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}\mathcal{C}^T = \sum_{i=1}^{k} \boldsymbol{c}_i\boldsymbol{c}_i^T \otimes \boldsymbol{\Sigma}_i$, where $\boldsymbol{c}_i = c_i\mathbf{1}$. Thus, $\mathcal{C}\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}\mathcal{C}^T = \sum_{i=1}^{k} c_i^2\mathbf{1}\mathbf{1}^T \otimes \boldsymbol{\Sigma}_i = \mathbf{1}\mathbf{1}^T \otimes \sum_{i=1}^{k} c_i^2\boldsymbol{\Sigma}_i$. Therefore,

$$\text{Var}(\mathbf{Y}) = \text{bdiag}(\sigma^2\mathbf{I}) + \mathbf{1}\mathbf{1}^T \otimes \sum_{i=1}^{k} c_1^2\boldsymbol{\Sigma}_i, \tag{17}$$

which is in quasi-Kronecker form and the inverse is easily calculable using equation (4) from [7].

This paper addresses the sequential nature of RMVs, one important step toward the improvement of QA/QC and IC for modern earthwork compaction applied to road construction. Improvement of the compaction process requires properly identifying soft, or weak spots and heterogeneous structures at the compaction site. Proper modeling of the sequential nature of data collection and fast and effective estimation procedures are

paramount to improving the compaction process. The sequential, spatial mixed-effects model provides the general flexibility required and the Sequential Backfitting Algorithm provides the necessary fast and effective estimation procedure to implement proper modeling of uncertainty in RMVs and improve quality of compaction and reduce compaction time. This assumes a static state of compaction of the site when the RMVs are collected. Proper modeling of the dynamic process of the earthwork compaction is still in progress.

Acknowledgements

The roller data was collected and provided by Dr. Mike Mooney (Colorado School of Mines, Golden, CO) as part of the NCHRP 21-09 project of the Transportation Research Board of The National Academies. This work is supported in part by the Swiss National Science Foundation through Project 200021-129782.

References

[1] L. A. Waller, B. P. Carlin, H. Xia, A. E. Gelfand, Hierarchical spatio-temporal mapping of disease rates, Journal of the American Statistical Association 92 (1997) 607–617.

[2] C. K. Wikle, L. M. Berliner, N. Cressie, Hierarchical Bayesian space-time models, Environmental and Ecological Statistics 5 (1998) 117–154.

[3] T. Bailey, W. Krzanowski, An overview of approaches to the analysis and modelling of multivariate geostatistical data, Mathematical Geosciences 44 (2012) 381–393.

[4] W. Christensen, S. Sain, Latent variable modeling for integrating output from multiple climate models, Mathematical Geosciences 44 (2012) 395–410.

[5] N. Cressie, C. K. Wikle, Statistics for Spatio-Temporal Data (Wiley Series in Probability and Statistics), Wiley, 2011.

[6] R. Furrer, S. R. Sain, Spatial model fitting for large datasets with applications to climate and microarray problems, Statistics and Computing 19 (2009) 113–128.

[7] D. K. Heersink, R. Furrer, On Moore–Penrose inverses of quasi-Kronecker structured matrices, Linear Algebra and its Applications 436 (2012) 561–570.

[8] H. Thurner, A. Sandström, Continuous compaction control, in: Proceedings of the European Workshop Compaction of Soils and Granular Materials, pp. 237–246.

[9] J. Scherocman, S. Rakowski, K. Uchiyama, Intelligent compaction, does it exist?, in: Proceedings of the 52 Annual Conference of the Canadian Technical Asphalt Association, Polyscience Publications Inc., 2007.

[10] M. A. Mooney, R. V. Rinehart, D. J. White, P. K. Vennapusa, N. W. Facas, O. M. Musimbi, Intelligent soil compaction systems: NCHRP project 21-09 final report, 2010.

[11] R. Anderegg, K. Kaufmann, Intelligent compaction with vibratory rollers: Feedback control systems in automatic compaction and compaction control, Transportation Research Record: Journal of the Transportation Research Board 1868 (2004) 124–134.

[12] M. Mooney, D. Adam, Vibratory roller integrated measurement of earthwork compaction: An overview, volume 307, ASCE, 2007, pp. 80–80.

[13] R. Rinehart, M. Mooney, Measurement depth of vibratory roller-measured soil stiffness, Géotechnique 59 (2008) 609–619.

[14] N. W. Facas, M. A. Mooney, Characterizing the precision uncertainty in vibratory roller measurement values, Journal of Testing and Evaluation 40 (2011) 1–9.

[15] W. Kröber, R. Floss, W. Wallrath, Dynamic soil stiffness as quality criterion for soil compaction, in: A. G. Correia, H. Brandl (Eds.), Geotechnics for Roads, Rail Tracks, and Earth Structures, Taylor & Francis, 2001, pp. 189–199.

[16] M. A. Mooney, R. V. Rinehart, Field monitoring of roller vibration during compaction of subgrade soil, Journal of Geotechnical and Geoenvironmental Engineering 133 (2007) 257–265.

[17] T. Yoo, E. Selig, Dynamics of vibratory-roller compaction, Journal of the Geotechnical Engineering Division 105 (1979) 1211–1231.

[18] D. Adam, F. Kopf, Operational devices for compaction optimization and quality control, in: Geotechnics in Pavement and Railway Design and Construction: Proceedings of the International Seminar on Geotechnics and Railway Design and Construction, Athens, Greece, 16 - 17 December 2004, IOS Press, 2004, pp. 97–106.

[19] P. J. van Susante, M. A. Mooney, Capturing nonlinear vibratory roller compactor behavior through lumped parameter modeling, Journal of Engineering Mechanics 134 (2008) 684–693.

[20] D. J. White, M. J. Thompson, Relationships between in situ and roller-integrated compaction measurements for granular soils, Journal of Geotechnical and Geoenvironmental Engineering 134 (2008) 1763–1770.

[21] M. N. M. van Lieshout, Sequential Spatial Processes For Image Analysis, in: V. Capasso (Ed.), Proceedings of Stereology and Image Analysis 2009, ESCULAPIO Co, 2009, pp. 1–6.

[22] D. K. Heersink, R. Furrer, Spatial analysis of modern soil compaction roller measurement values, Procedia Environmental Sciences 7 (2011) 8 – 13. Spatial Statistics 2011: Mapping Global Change.

[23] N. Cressie, Statistics for spatial data, Wiley series in probability and mathematical statistics: Applied probability and statistics, J. Wiley, 1993.

[24] D. Harville, Matrix Algebra From a Statistician's Perspective, Springer, Heidelberg, 1997.

[25] A. Buja, T. Hastie, R. Tibshirani, Linear smoothers and additive models, The Annals of Statistics 17 (1989) 453–510.

[26] L. Breiman, J. H. Friedman, Estimating optimal transformations for multiple regression and correlation, Journal of the American Statistical Association 80 (1985) 580–598.

[27] M. Genton, Variogram Fitting by Generalized Least Squares Using an Explicit Formula for the Covariance Structure, Mathematical Geology 30 (1998) 323–345.

[28] N. W. Facas, Variogram Properties and Anisotropy in the Spatial Distribution of Roller-Measured Soil Stiffness, Master's thesis, Colorado School of Mines, 2009.

[29] S. Wood, Generalized Additive Models: An Introduction with R, Texts in Statistical Science, Chapman & Hall/CRC, 2006.

Paper IV

Intelligent Compaction and Quality Assurance of Roller Measurement Values utilizing Backfitting and Multiresolution Scale Space Analysis

Daniel K. Heersink, Reinhard Furrer, & Mike A. Mooney

Paper submitted to *Mathematical Geosciences*.

Mathematical Geosciences manuscript No.
(will be inserted by the editor)

Intelligent Compaction and Quality Assurance of Roller Measurement Values utilizing Backfitting and Multiresolution Scale Space Analysis

Daniel K. Heersink · Reinhard Furrer · Mike A. Mooney

Received: date / Accepted: date

Abstract Modern earthwork compaction rollers collect a virtually continuous flow of location and compaction information as they traverse a compaction site. These roller measurement values present a challenging spatio-temporal statistical problem that requires careful implementation of a proper stochastic model and estimation procedure, but also hold the potential to improve intelligent compaction and quality assurance of the compaction process. Modern earthwork compaction for road construction is a sequential process involving several layers of material. The sequential nature of these roller measurement values is modeled as a sequential, spatial mixed-effects model and modeling terms are estimated using a sequential, spatial backfitting routine. Bayesian critical regions of high and low compaction areas are produced from the estimated fields from this backfitting procedure using a multiresolution scale space analysis. This regime of stochastic model, estimation procedure and image analysis is computationally feasible for on board computers and is proposed as a viable solution to improved intelligent compaction and quality assurance of the compaction process.

Keywords Spatio-temporal data · sequential modeling · image analysis · Bayesian statistics

D.K. Heersink
Institute of Mathematics, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland

R. Furrer
Institute of Mathematics, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland
Tel.: +41-(0)44-6355843
Fax: +41-(0)44-6355705
E-mail: reinhard.furrer@math.uzh.ch

M.A. Mooney
Civil and Environmental Engineering Department, Colorado School of Mines, 1500 Illinois St., Golden, CO 80401 USA

1 Introduction

Current road construction practice utilizes modern earthwork compaction rollers to compact several layers of material during the road building process. A vibrating drum with a diameter of approximately 1m traverses the compaction site in several adjacent driving lanes. The number of lanes depends on the width of the road under construction and is usually 5 to 7. These rollers are outfitted with a sensor and GPS unit to collect compaction and location data as they operate. This data, termed the roller measurement value (RMV), can be used to improve the quality of the compaction process. This is achieved by improving the identification of weak, or soft spots, and by ensuring a homogeneous compaction of the site material.

The first modern earthwork compaction rollers designed for continuous compaction control (CCC) were used in practice starting in the 1970s in the European community. CCC is a method of documenting compaction and is used to achieve homogeneous compaction in a minimum time (Thurner and Sandström, 2000). Each roller manufacturer has developed a proprietary measurement of soil stiffness used for CCC. Current use of RMVs is the identification of potential areas of soft, or weak, spots. Acceptance of these areas is based on the weak spots meeting prespecified criteria (Mooney et al., 2010).

Each RMV reflects an aggregated volume of soil measuring a bulb extending to a depth of approximately 1m with a diameter of 0.5 to 0.6m (Facas, 2009). Typical construction practice is to compact in segments of road 10 to 15m wide and 50 to 100m long. See Fig. 1 for a representative roller manufactured by Ammann.

Fig. 1 Ammann roller at work.

These RMVs are calibrated to the true underlying soil stiffness and to spot testing procedures at the beginning of the compaction process. They are thus directly proportional to the stiffness of the underlying material and the bias of the RMVs cannot be determined. Depending on the manufacturer, the RMV can be a direct representation of the underlying stiffness in the form of

a spring constant or can be an indirect representation through a measure of nonlinearity.

Previous geostatistical investigations of RMVs focus on identifying sources of uncertainty and anisotropy (Facas, 2009; Facas et al., 2010; Facas and Mooney, 2011). Facas and Mooney (2011) identify two values of uncertainty in RMVs. The uncertainty in a single RMV is termed σ_1 and the uncertainty in the difference between two RMVs collected at the same spatial location is σ_2. Under the assumption of normality, $\sigma_2 = \sqrt{2}\sigma_1$. Two methods are used to empirically determine σ_2. The first method computes the standard deviation of the difference vector of two passes through the compaction site and is denoted $\widehat{\sigma}_2^D$. The second is to estimate the nugget from an empirical semivariogram of the data and is denoted $\widehat{\sigma}_2^N$. Semivariogram analysis has been implemented on RMV data in other contexts by Facas and Mooney (2011); Petersen et al. (2007); White and Thompson (2008).

Rudimentary intelligent compaction (IC) technology was first available in the late 1990s. IC is an automated system that adjusts roller operation parameters for optimal compaction based on CCC data (Scherocman et al., 2007). IC is a development aimed at improving quality assurance (QA) of the compaction process. This rudimentary IC uses the raw, calibrated data from the sensor to identify regions of the compaction site that require closer inspection and spot testing. Utilization of the inherent spatial uncertainty in modeling and estimation of this data, as done here, will lead to improved IC and QA.

RMVs present a challenging spatio-temporal modeling problem. The sequential construction process introduces anisotropy into the data (Facas et al., 2010). They exhibit multiple scales of variation and are nonstationary. In order to improve QA and rudimentary IC, estimation and decision making procedures must be robust and computationally efficient. These procedures must also be implemented on a standard laptop that can be carried on-board the roller.

In Sect. 2 the two datasets under investigation are detailed. A sequential, spatial mixed-effects model is proposed to adequately address the modeling concerns of RMVs in Sect. 3.1 and a sequential, spatial backfitting routine for estimation of model parameters are proposed in Sect. 3.2. The estimates are then analyzed using a multiresolution scale space analysis in Sect. 3.3. Section 5 presents concluding remarks and directions for further research.

2 Minnesota and Florida Test Beds

As part of the NCHRP 21-09 project of the Transportation Research Board of The National Academies, Dr. Mike Mooney and his team collected RMV data from a test bed using a smooth, vibrating drum roller manufactured by Ammann (Mooney et al., 2010). The test bed lies along a stretch of road adjacent to Interstate 94 ($45°15'45'', -93°42'37''$). The test bed is approximately 300 meters in length by 15 meters in width and divided into two cells, labeled 27 and 28. Measurements from the roller of the subsurface, subgrade, and base

layers of the new road under construction were recorded. Figure 2 is a plot of the data from both cells and all three layers.

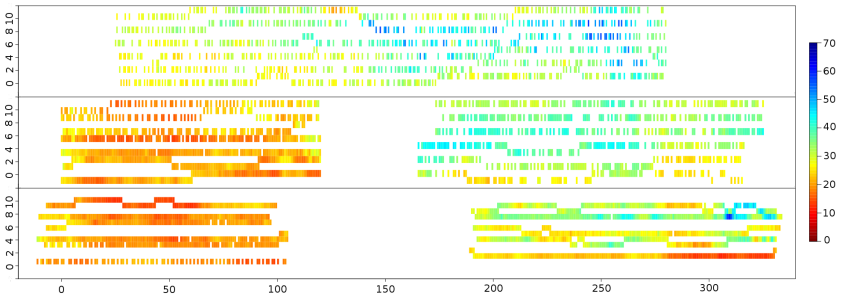

Fig. 2 Plot of subsurface (top), subgrade (middle), and base (bottom) data (in units of MN/m) from cells 27 (right) and 28 (left) from the test bed in Albertville, MN, USA. x-direction is the direction of driving.

Data from each layer was collected separately for each of the two cells, except for the subsurface layer where the roller drove continuously over both cells. Analysis was performed on each cell separately as the standard construction procedure focuses on one cell at a time. The subsurface layer was thus split into two cells at the cell boundary.

The subsurface layer is the existing material (clay) at the construction site. The subgrade layer consists of moisture conditioned clay and the base layer consists of a granular composite material (Mooney et al., 2010). The roller traversed the construction site in five to seven lanes, with the coordinates calculated by on-board GPS. Thus, the locations of the observation vector for each layer are unique.

A thorough investigation of the role driving direction of modern compaction rollers and location of the compaction sensor play in the measurement of RMVs was undertaken at a test bed in Florida. The compaction roller was outfitted with two sensors, one on the right, 0.41m from the end, and one on the left, 0.17m from the end. The location of the sensor is not of importance for the current presentation, so only data from the left sensor will be utilized.

This test bed is an atypically large, atypically compacted site. The compaction roller was driven across the compaction site in both the typical driving direction and also in the transverse direction. The compaction site is approximately square, 30m on each side. As this investigation focused on the effects of driving direction, only data for one layer of compaction was collected. Figure 3 is a plot of the data for the typical driving direction. This subset, left sensor and typical driving direction, of the data will be used in this paper.

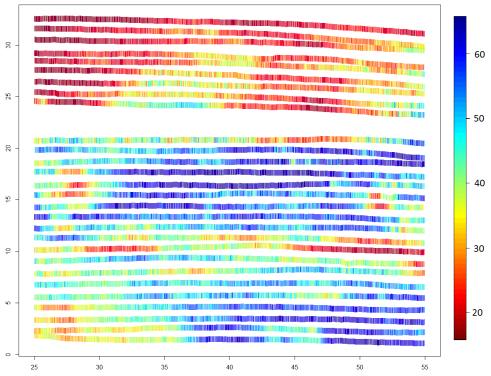

Fig. 3 Plot of the subset of the Florida test bed dataset (in units of MN/m) consisting of data collected from the left sensor while the compaction roller was traversing the compaction site in the typical driving direction.

3 Proposed Methodology for Quality Assurance and Intelligent Compaction

To improve both QA and IC, a sequential stochastic model of the construction site must be built that properly models anisotropy, multiple scales of variation, and nonstationarity. Proper estimation procedures for the parameters in these models must also be implemented. Once estimates have been made, a decision must be made of where, if anywhere, more compaction is needed. All of this modeling, estimation, and decision making must be robust (in the sense that it does not require an expert user), computationally efficient, and implementable on a standard on-board roller computer.

A sequential, spatial mixed-effects model was introduced by Heersink and Furrer (2013) to handle the multiple scales of variation inherent in RMVs. A sequential modeling approach is utilized for the sequential nature of RMVs. Heersink and Furrer (2013) also develop a sequential, spatial backfitting routine for estimation of the parameters of such sequential models. This sequential modeling and estimation procedure can be implemented into a more comprehensive program improving QA of the compaction process and also improving IC.

Holmström et al. (2011) developed multiresolution scale space analysis methods for image processing using a Bayesian framework. This methodology was demonstrated using several test images and is applied to climate change prediction fields. The method can also be used to analyze images of RMV estimates to identify weak, or soft, spots and large variations in the compaction area, that is QA and IC.

3.1 Sequential, Spatial Mixed-Effects Model

Heersink and Furrer (2013) introduced a sequential, spatial mixed-effects model for complex spatial structures and apply that model to the Minnesota test bed dataset detailed in Section 2. Let $\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\alpha} + \boldsymbol{\varepsilon}$ be the additive decomposition of the RMV data $\boldsymbol{y}$ into large scale trend $\mathbf{X}\boldsymbol{\beta}$ (fixed), small scale trend $\boldsymbol{\alpha}$ (random) and noise $\boldsymbol{\varepsilon}$. This is the standard spatial mixed-effects model.

RMV data is collected sequentially, that is, there is a new observation vector $\boldsymbol{y}_t$ for each layer of the compaction process. The sequential, spatial mixed-effects model thus adds new small scale trends for each additional layer:

$$\begin{aligned} \boldsymbol{y}_1 &= \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\alpha}_1 + \boldsymbol{\varepsilon}_1, \\ \boldsymbol{y}_2 &= \mathbf{X}_2\boldsymbol{\beta}_2 + c\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2 + \boldsymbol{\varepsilon}_2, \\ \boldsymbol{y}_3 &= \mathbf{X}_3\boldsymbol{\beta}_3 + c^2\boldsymbol{\alpha}_1 + c\boldsymbol{\alpha}_2 + \boldsymbol{\alpha}_3 + \boldsymbol{\varepsilon}_3. \end{aligned} \tag{1}$$

Here, $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3$ are independent Gaussian random fields with covariance matrices $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \boldsymbol{\Sigma}_3$, and $\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \boldsymbol{\varepsilon}_3$ are uncorrelated white noise processes with variances $\sigma_1^2, \sigma_2^2, \sigma_3^2$, mutually independent of the $\boldsymbol{\alpha}_i$s.

The parameter c governs how much of previous layers are "seen" by the current observation. As the roller measures to a depth of several layers of compaction material, the current measurement is an aggregate measurement of the compaction of several layers of material. The model is thus of a spatial "autoregressive" type. The $\mathbf{X}_t$ matrices are fixed effects components consisting of columns of powers of x and y coordinates, and potentially other roller parameters like amplitude and frequency.

Computation time can be further reduced by mapping the observations to a regular grid. This is accomplished by implementing a state-space formulation to handle the unique data observation locations by introducing an operator matrix into (1) that maps observed locations to a regular grid:

$$\begin{aligned} \boldsymbol{y}_1 &= \mathbf{H}_1\mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{H}_1\boldsymbol{\alpha}_1 + \boldsymbol{\varepsilon}_1, \\ \boldsymbol{y}_2 &= \mathbf{H}_2\mathbf{X}_2\boldsymbol{\beta}_2 + c\mathbf{H}_2\boldsymbol{\alpha}_1 + \mathbf{H}_2\boldsymbol{\alpha}_2 + \boldsymbol{\varepsilon}_2, \\ \boldsymbol{y}_3 &= \mathbf{H}_3\mathbf{X}_3\boldsymbol{\beta}_3 + c^2\mathbf{H}_3\boldsymbol{\alpha}_1 + c\mathbf{H}_3\boldsymbol{\alpha}_2 + \mathbf{H}_3\boldsymbol{\alpha}_3 + \boldsymbol{\varepsilon}_3. \end{aligned} \tag{2}$$

Here, $\boldsymbol{\alpha}_t, \boldsymbol{\Sigma}_t, \boldsymbol{\varepsilon}_t, \sigma_t^2$, $t = 1, 2, 3$ are as in (1), with the exception that the $\boldsymbol{\alpha}_t$ terms are on a regular grid.

For the purposes of prediction, the driving associated fixed effects are eliminated from the $\boldsymbol{\beta}_t$ terms. Explicitly, we could write the fixed effects as $\mathbf{X}_t'\boldsymbol{\beta}_t' + \mathbf{X}_t\boldsymbol{\beta}_t$, where $\mathbf{X}_t'\boldsymbol{\beta}_t'$ corresponds to the driving associated fixed effects. Thus we would then have that

$$\boldsymbol{y}_t = \mathbf{X}_t'\boldsymbol{\beta}_t' + \mathbf{H}_t(\mathbf{X}_t\boldsymbol{\beta}_t + c^2\boldsymbol{\alpha}_{t-2}I_{(t>2)} + c\boldsymbol{\alpha}_{t-1}I_{(t>1)} + \boldsymbol{\alpha}_t) + \boldsymbol{\varepsilon}_t,$$

$t = 1, 2, 3$, where $I_{(t>\tau)}$ is the indicator function that is one when $t > \tau$ and zero otherwise.

This sequential modeling framework has several advantages. First, the sequential nature of this modeling approach decomposes the measurements into

individual contributions from each layer of the compaction process. This decomposition has the potential to better inform the roller operator on the compaction state of the current layer of compaction as opposed to providing an aggregate measure of compaction of several layers of material.

Second, for a viable modeling and estimation procedure, computation time must be minimized. To aid in computational time, a spherical covariance structure is utilized as this produces sparse matrices. Empirical semivariograms of RMV data suggest the spherical covariance structure is adequate and an infinite range covariance structure is not necessary.

3.2 Backfitting of Roller Measurement Values

The backfitting algorithm has been employed on a wide range of additive models (Breiman and Friedman, 1985; Buja et al., 1989; Furrer and Sain, 2009). The basic model for the backfitting algorithm is $\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\alpha} + \boldsymbol{\varepsilon}$, where $\mathrm{Var}(\boldsymbol{\alpha}) = \boldsymbol{\Sigma}(\boldsymbol{\theta})$ and $\mathrm{Var}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}$, $\boldsymbol{\alpha}$ independent of $\boldsymbol{\varepsilon}$. Here, $\boldsymbol{\theta}$ is the parameter vector for the covariance matrix of $\boldsymbol{\alpha}$. If the parameters $\boldsymbol{\theta}$ and σ are known, the backfitting algorithm then iteratively estimates the fixed effects using generalized least squares and the spatial effects using spatial smoothing:

$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}^\mathsf{T}\boldsymbol{\Sigma}_{\boldsymbol{y}}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^\mathsf{T}\boldsymbol{\Sigma}_{\boldsymbol{y}}^{-1}\boldsymbol{y} \quad \text{(generalized least-squares estimator)},$$
$$\widehat{\boldsymbol{\alpha}} = \boldsymbol{\Sigma}(\boldsymbol{\theta})\boldsymbol{\Sigma}_{\boldsymbol{y}}^{-1}\left(\boldsymbol{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}\right) \quad \text{(spatial smoothing)},$$

where $\boldsymbol{\Sigma}_{\boldsymbol{y}} = \mathrm{Var}(\boldsymbol{y}) = \boldsymbol{\Sigma}(\boldsymbol{\theta}) + \sigma^2\mathbf{I}$, the covariance matrix of the observations. The spatial backfitting algorithm produces new iterative estimates until they converge, that is the estimates no longer differ with each iteration, up to a small number.

Heersink and Furrer (2013) extend the backfitting algorithm to accommodate multiple spatial processes and the sequential, spatial mixed-effects model. This extension of the backfitting algorithm, termed the Sequential Backfitting Algorithm, sequentially estimates covariance parameters of several spatial processes through multiple spatial smoothing steps. Formal proof of convergence of the algorithm, as well as a simulation study demonstrating convergence of the algorithm in practice are provided in Heersink and Furrer (2013).

The algorithm is:

Sequential Backfitting Algorithm

[0] For $t = 1$ to p
[1] Let $\widehat{\boldsymbol{\alpha}}_t^{(0)}$ be an initial guess and put $j = 0$
[2] $j = j + 1$
[3] $\widehat{\boldsymbol{\beta}}_t^{(j)} = \left(\mathbf{X}_t^T\mathbf{X}_t\right)^{-1}\mathbf{X}_t^T\left(\boldsymbol{y}_t - \sum_{i=1}^{t} c^{t-i}\widehat{\boldsymbol{\alpha}}_i^{(j-1)}\right)$

[4] Estimate covariance parameters to get $\widehat{\sigma}_t^{2,(j)}$ and $\widehat{\boldsymbol{\Sigma}}_t^{(j)}$, then put
$\widehat{\boldsymbol{\alpha}}_t^{(j)} = \widehat{\boldsymbol{\Sigma}}_t^{(j)} \left(\widehat{\sigma}_t^{2,(j)}\mathbf{I} + \widehat{\boldsymbol{\Sigma}}_t^{(j)}\right)^{-1} \left(\boldsymbol{y}_t - \mathbf{X}_t\widehat{\boldsymbol{\beta}}_t^{(j)} - \sum_{\ell=1}^{t-1} c^{t-\ell}\widehat{\boldsymbol{\alpha}}_\ell^{(j)}\right)$

[5] Repeat [3] to [5] until convergence.

3.3 Multiresolution Scale Space Analysis

To investigate significant features of an image at multiple scales, Holmström et al. (2011) developed a Bayesian framework of image processing at multiple resolutions. This multiresolution scale space analysis identifies credible regions of the image that are statistically significantly more positive or negative than surrounding regions. Multiresolution scale space analysis is a method of simultaneously smoothing an input, such as data or an image, at several levels. Each smooth of the input provides a different scale of information. A brief summary of scale space multiresolution analysis, as detailed in Holmström et al. (2011), follows.

This scale space multiresolution procedure is accomplished in three steps. First, the signal, or image $\boldsymbol{y}$, is synthesized or reconstructed using a Bayesian framework. Let $\mathbf{S}_\lambda$ be a smoothing operator represented by a square matrix, where λ is the parameter governing the level of smoothing. The image is then smoothed for a series of λ values $0 = \lambda_1 < \lambda_2 < \cdots < \lambda_{L-1} < \lambda_L = \infty$.

Second, the differences of this series of images are used to form scale-dependent detail components of the reconstructed image. The image is reconstructed as

$$\mathbf{X}_t\boldsymbol{\beta}_t + \boldsymbol{\alpha}_t = \sum_{i=1}^{L-1}(\mathbf{S}_{\lambda_i} - \mathbf{S}_{\lambda_i+1})(\mathbf{X}_t\boldsymbol{\beta}_t + \boldsymbol{\alpha}_t) + \mathbf{S}_{\lambda_L}(\mathbf{X}_t\boldsymbol{\beta}_t + \boldsymbol{\alpha}_t) \equiv \sum_{i=1}^{L-1}\boldsymbol{z}_i + \boldsymbol{z}_L,$$

where $\boldsymbol{z}_i$ is the difference of two consecutive smooths and $\boldsymbol{z}_L = \mathbf{S}_\infty(\mathbf{X}_t\boldsymbol{\beta}_t + \boldsymbol{\alpha}_t)$, the overall mean value of the image as defined by the choice of $\mathbf{S}_\lambda$. Each of these $\boldsymbol{z}_i$s is now a scale-dependent detail component image of the original image.

Finally, each scale-dependent detail component image is analyzed to create credibility maps. A posterior probability threshold $0 < \alpha < 1$ is chosen. For this analysis, the typical value of $\alpha = 0.95$ was chosen. The credibility map is then created by coloring location (x, y) of image $\boldsymbol{z}_i$, $z_i(x, y)$, blue if $P(z_i(x, y) > 0|\mathbf{X}_t\boldsymbol{\beta}_t + \boldsymbol{\alpha}_t) > \alpha$, red if $P(z_i(x, y) < 0|\mathbf{X}_t\boldsymbol{\beta}_t + \boldsymbol{\alpha}_t) > \alpha$, and gray otherwise. This inference is done jointly for all locations, instead of each location separately.

The smoothing levels must be chosen with some care as to adequately decompose the image into its scale-dependent detail component images. Components of the image on a scale not chosen for smoothing could be lost. Also,

the same components could be found on multiple adjacent smoothing scales if the selection of λs is too fine.

The smoothing levels used in this multiresolution scale space analysis can be optimized using an objective function. For $L = 4$, for example, the objective function chosen for minimization for the parameter sequence $[0, \lambda_2, \lambda_3, \infty]$ is

$$G(\lambda_2, \lambda_3) = \sum_{\{i,j=1,2,3|i<j\}} \frac{|\widetilde{\boldsymbol{\gamma}}_i{}^{\mathsf{T}} \widetilde{\boldsymbol{\gamma}}_j|}{\|\widetilde{\boldsymbol{\gamma}}_i\| \|\widetilde{\boldsymbol{\gamma}}_j\|}, \tag{3}$$

where the $\widetilde{\boldsymbol{\gamma}}_i$s are the signal-dependent tapering functions defining the $\boldsymbol{z}_i$s through $\mathbf{X}_t\boldsymbol{\beta}_t + \boldsymbol{\alpha}_t$ and the eigendecomposition of $\mathbf{S}_\lambda$. The minimization is then done with respect to λ_2 and λ_3. The objective function can be extended to any number of smoothing levels, but optimization becomes difficult in higher dimensions.

3.4 Minnesota Test Bed Analysis

For the Minnesota test bed dataset, let $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3$ in (2) be full rank matrices of covariates consisting of an intercept, centered and scaled (x, y)-coordinates of the RMV, centered and scaled x-coordinates of the second and third power, and the driving direction of the roller (1 for right-to-left, 0 for left-to-right). Only higher order powers of the x-coordinate were used due to the scale difference in the two directions. $\mathbf{H}_1, \mathbf{H}_2, \mathbf{H}_3$ are matrices mapping coordinates of the process level to observed locations.

Centering and scaling of the coordinates addresses the range anisotropy in RMVs, see Facas et al. (2010) for discussion of anisotropy in RMVs. This centering and scaling of the coordinate system is a simple, ad hoc transformation of the coordinate system to account for the range anisotropy (Zimmerman, 1993).

The process level estimates of the compaction process were obtained from the Sequential Backfitting Algorithm (Heersink and Furrer, 2013), as described above. These estimates must then be analyzed to assess the quality of the compaction for each layer and cell of the compaction site.

In practice, there is a site specific threshold value of compaction required for the layer to be deemed sufficiently compacted. This threshold value can be dependent on the current material of compaction. For this analysis, a threshold value of 20 was used for all cells and layers for demonstration. The threshold value was subtracted from all images such that values less than zero represent areas that are too soft. Theoretically, the images are invariant to an additive constant such as this thresholding procedure, that is, the thresholding does not change the range of the estimates thus the images are produced on the same color scale, irrespective of the threshold value. The thresholding was done for a practical advantage of more easily identifying soft areas.

The objective function (3) was minimized for all three layers of each cell of the compaction site. Similar optimum values of the two smoothing parameters

were found for all layers and both cells. See Fig. 4 for a representative plot of (3) for the subsurface layer of cell 27. For all layers of compaction and both cells, $\lambda_2 \in (0, 4)$ and $\lambda_3 \in (290, 1700)$, (Table 1).

Table 1 Minimum values (λ_2, λ_3) of (3) for subsurface, subgrade, and base layers of cell 27 (left) and cell 28 (right).

	Cell 27	Cell 28
subsurface	$(1.09, 1079.9)$	$(1.89, 294.5)$
subgrade	$(0.54, 1693.2)$	$(0.52, 378.1)$
base	$(1.09, 623.2)$	$(3.62, 309.6)$

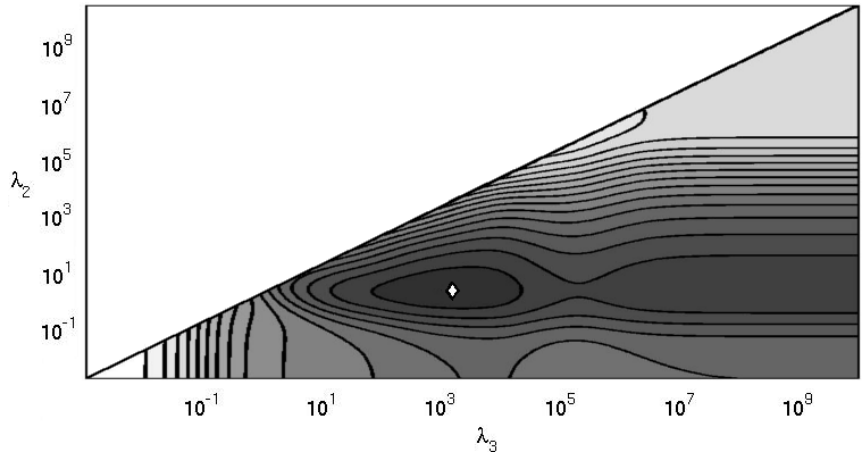

Fig. 4 Minimization of the objective function (3) for the Minnesota test bed, subsurface layer of cell 27. The minimum point is identified by a white diamond. The minimization for other layer and cell combinations produces similar figures.

Multiresolution scale space analysis was then applied to the six images obtained from the backfitting algorithm to identify which features in the images are real features (in the Bayesian confidence region sense) and which are artifacts of random variation. Each image created from the backfitting algorithm consists of the process level estimates of a specific layer and cell: $\mathbf{X}_t\widehat{\boldsymbol{\beta}}_t + \widehat{\boldsymbol{\alpha}}_t$.

As the true distribution of the estimates can be calculated in a straightforward manner, a true Bayesian framework with priors and posteriors was not needed. Instead, 500 samples were drawn from the "posterior" of $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, $\boldsymbol{\beta}_3$, $\boldsymbol{\alpha}_1$, $\boldsymbol{\alpha}_2$, and $\boldsymbol{\alpha}_3$. Here, the posterior terminology is used to draw equivalencies to a true Bayesian model. The distributions used are multivariate Gaussian with a mean and covariance matrix of the estimates. That is, samples of the

$\boldsymbol{\beta}_t$s were drawn from

$$\boldsymbol{\beta}_t \sim \mathcal{N}\left(\widehat{\boldsymbol{\beta}}_t, \left(\mathbf{X}_t{}^{\mathsf{T}}\mathbf{H}_t{}^{\mathsf{T}}\widehat{\boldsymbol{\Sigma}}_{y,t}^{-1}\mathbf{H}_t\mathbf{X}_t\right)^{-1}\right) \tag{4}$$

and samples of the $\boldsymbol{\alpha}_t$s were drawn from

$$\boldsymbol{\alpha}_t \sim \mathcal{N}\left(\widehat{\boldsymbol{\alpha}}_t, \widehat{\boldsymbol{\Sigma}}_t\mathbf{H}_t{}^{\mathsf{T}}\widehat{\boldsymbol{\Sigma}}_{\alpha,t}\mathbf{H}_t\widehat{\boldsymbol{\Sigma}}_t\right), \tag{5}$$

where,

$$\widehat{\boldsymbol{\Sigma}}_{\alpha,t} = \left(\widehat{\boldsymbol{\Sigma}}_{y,t}^{-1} - \widehat{\boldsymbol{\Sigma}}_{y,t}^{-1}\mathbf{H}_t\mathbf{X}_t\left(\mathbf{X}_t{}^{\mathsf{T}}\mathbf{H}_t{}^{\mathsf{T}}\widehat{\boldsymbol{\Sigma}}_{y,t}^{-1}\mathbf{H}_t\mathbf{X}_t\right)^{-1}\mathbf{X}_t{}^{\mathsf{T}}\mathbf{H}_t{}^{\mathsf{T}}\widehat{\boldsymbol{\Sigma}}_{y,t}^{-1}\right), \text{ and}$$
$$\widehat{\boldsymbol{\Sigma}}_{y,t} = \mathbf{H}_t\widehat{\boldsymbol{\Sigma}}_t\mathbf{H}_t{}^{\mathsf{T}} + \widehat{\sigma}_t^2\mathbf{I}.$$

Strictly speaking, $\widehat{\boldsymbol{\beta}}_t$ and $\widehat{\boldsymbol{\alpha}}_t$ are not independent. If we were to use the dependency, the standard deviation of the sum would decrease. Thus we overestimate the variabilty, but hardly any effect can be seen in the final red/blue image.

Following the minimization of (3), the set of smoothing levels used was $[4, 1000, \infty]$. These smoothing levels were chosen for all compaction layers and both cells to maintain consistency across all images. The smallest smoothing level, $\lambda = 4$, was chosen to provide credible regions on the scale of the roller width and corresponds to a small scale structure on the order of 2-5m. The $\lambda = 1000$ value was chosen to identify large scale trends in the compaction region and corresponds to an overall mean and large scale mean structure on the order of 75m. Both of these correspond to the optimum values identified by the objective function minimization.

Figure 5 depicts the results of this multiresolution scale space analysis for the subsurface, subgrade, and base layers of cell 27. Figure 6 depicts the analysis for cell 28.

3.5 Florida Test Bed Analysis

For the Florida test bed dataset, there is only one compaction layer and a sequential, spatial model is unnecessary. Instead, the standard spatial mixed-effects model can be used. Thus, the Florida test bed data can be decomposed as $\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\alpha} + \boldsymbol{\varepsilon}$. Since the compaction site for this test bed is square, let $\mathbf{X}$ be the matrix of fixed effects defined by all polynomial terms of the x- and y-direction up to fifth order, including cross terms. Let $\boldsymbol{\alpha}$ be a spatially correlated process and $\boldsymbol{\varepsilon}$ be a noise process as for the Minnesota test bed dataset.

As this dataset comes from a compaction site that is already square, a transformation of the coordinates does not make sense. Therefore, an additional parameter is introduced into the covariance structure of the $\boldsymbol{\alpha}$ term to account for the inherent anisotropy in RMV data.

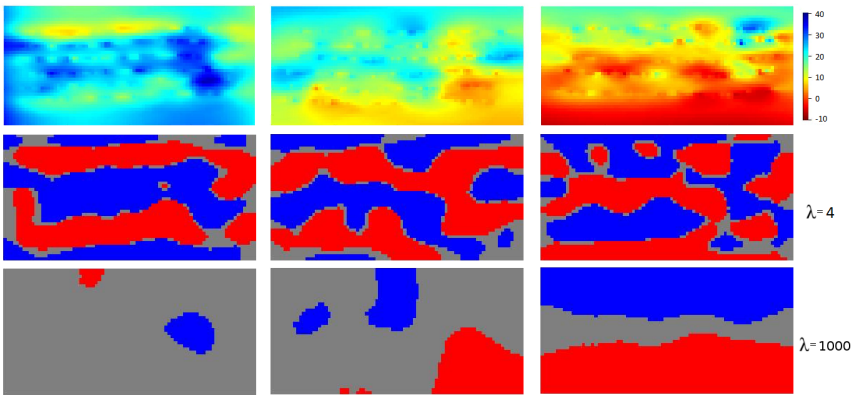

Fig. 5 Plot of estimated subsurface (left), subgrade (middle), and base (right) for cell 27 from the test bed in Albertville, MN, USA (top) on the regular grid. Credibility plots from a multiresolution scale space analysis are depicted in the bottom two plots for different values of λ. Red corresponds to softer RMVs. All credibility plots for $\lambda = \infty$ are solid blue, indicating an overall sufficient compaction has been attained with the given threshold.

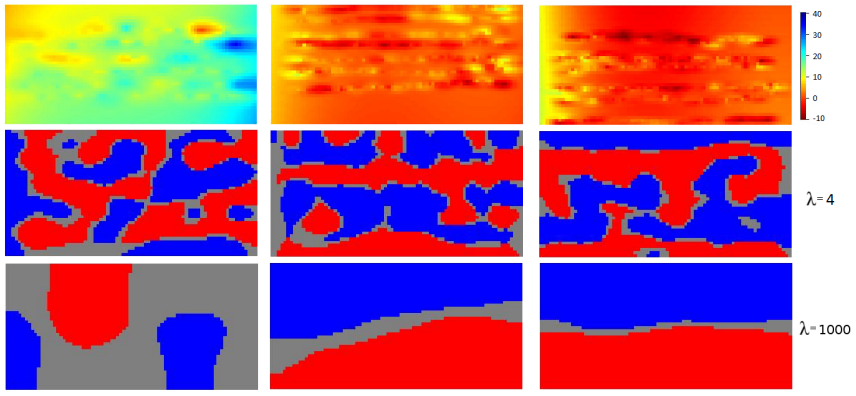

Fig. 6 Plot of estimated subsurface (left), subgrade (middle), and base (right) for cell 28 from the test bed in Albertville, MN, USA (top) on the regular grid. Credibility plots from a multiresolution scale space analysis are depicted in the bottom two plots for different values of λ. Red corresponds to softer RMVs. All credibility plots for $\lambda = \infty$ are solid blue, indicating an overall sufficient compaction has been attained with the given threshold.

To aid in computation time, a random sample of 2500 data points from the selected subset of the data was used. The analysis was repeated with a new random sample of 2500 data points and no appreciable differences were detected.

A state space formulation is also used for this dataset to map observed locations to a regular grid. The grid chosen for this dataset is a 70×70 square encompassing all locations. Let $\mathbf{H}$ be the matrix that maps the observed locations to this grid. The data is thus modeled as $\boldsymbol{y} = \mathbf{H}\mathbf{X}\boldsymbol{\beta} + \mathbf{H}\boldsymbol{\alpha} + \boldsymbol{\varepsilon}$. The

Sequential Backfitting Algorithm can then be utilized to estimate parameters of the model by setting $p = 1$. Note that this is equivalent to the classical backfitting approach.

The objective function (3) was minimized for the Florida test bed compaction site as well. The optimum minimum values identified were 0 and 1800. This suggests only one λ value between 0 and ∞ is necessary. See Fig. 7 for a plot of (3) for the Florida test bed dataset.

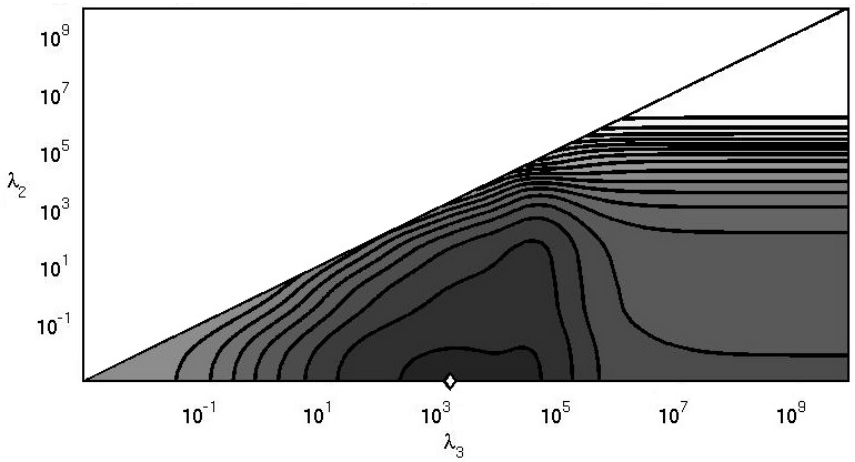

Fig. 7 Minimization of the objective function (3) for the Florida test bed. The minimum point is identified by a white diamond.

Following the minimization of (3), and following the advice of Holmström et al. (2011), the set of smoothing levels used was $[0, 900, 1800, \infty]$. See Fig. 8 for the results of the multiresolution scale space analysis. The smallest smoothing level again provides credible regions on the scale of the roller. The $\lambda = 1800$ smoothing level describes large scale mean structures and the $\lambda = 900$ smoothing level provides intermediate characteristics of the compaction site. Again, the $\lambda = \infty$ smoothing level plot is solid blue and omitted.

Multiresolution scale space analysis was then applied to the process level estimates obtained from the backfitting algorithm. The image consists of the process level estimates $\mathbf{X}\widehat{\boldsymbol{\beta}} + \widehat{\boldsymbol{\alpha}}$.

Again, the true distribution of the estimates can be calculated in a straightforward manner. 500 samples were drawn from the "posterior" of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$. The distributions of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ are (4) and (5) respectively, where there is no need for the subscript t as $p = 1$.

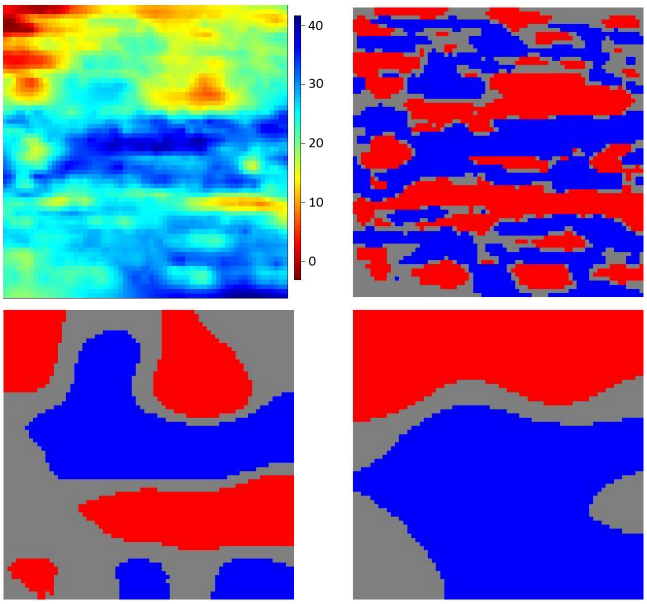

Fig. 8 Plot of estimated compaction from the Florida test bed (top left). Credibility plots from a multiresolution scale space analysis are depicted in the remaining three plots for $\lambda = 0$ (top right), $\lambda = 900$ (bottom left), and $\lambda = 1800$ (bottom right). Red corresponds to softer RMVs. The credibility plot for $\lambda = \infty$ is solid blue, indicating an overall sufficient compaction has been attained with the given threshold.

4 Discussion

The blue color in the images generated by the multiresolution scale space analysis indicates credible regions of sufficient compaction. The red color indicates credible regions where a sufficient level of compaction is suspect. Figures 5, 6, and 8 detail credible regions of heterogeneity of the compaction region. These areas could have been compacted more to achieve a more homogeneous compaction and improve QA of the construction.

4.1 Minnesota Test Bed

The banding nature of the estimates obtained from the sequential modeling and backfitting procedure found in Figs. 5 and 6 are a product of the size of the grid used for the process level estimates. For the subgrade, and especially the subsurface layers of the compaction site, the quantity of data is much less than that for the base layer of compaction. This relative lack of data causes an overfitting of the data, resulting in the banding structure evident in Figs. 5 and 6. This banding could be mitigated by choosing a coarser grid for the compaction layers with fewer data locations, but the authors chose to use one

grid for all layers for consistency. As this is data from a test site, there is more variation in the number of locations measured in each layer of compaction. A more uniform data collection for each layer would be expected in practice.

The $\lambda = 4$ images are most informative for identifying areas requiring more compaction as they identify regions on the same scale as the roller. The larger λ values provide a more general idea of the compaction of the entire cell by depicting a gradient of hard to soft values.

Cell 28 is consistently softer than cell 27 through all layers of compaction. This is an example of heterogeneity in the pre-existing material at the construction site, as represented by the subsurface layer. The features of the subsurface layer are inherited by subsequent layers of compaction. The red regions on the generated images, especially at the $\lambda = 4$ smoothing level, can be seen in all three layers. This is an expected result as each layer of compaction is on the order of 20cm thick and the roller measures to a depth on the order of 1m (Facas, 2009).

Cell 27 displays a gradient of more compact to less compact material from top to bottom in the subgrade layer. This gradient is still evident in the base layer. For cell 28, a gradient of more compact to less compact to more compact material from left to right in the subsurface layer is evident. This gradient changes to a more compact to less compact gradient moving top to bottom in the subgrade layer and remains through the base layer.

4.2 Florida Test Bed

The $\lambda = 0, 900$ images are most informative for identifying areas of the compaction site that could potentially need more compaction. The $\lambda = 0$ image identifies these regions on the scale of the roller, while the $\lambda = 900$ image identifies these regions on the scale of several widths of the roller.

The lower half of the compaction site is stiffer than the upper half, as can be seen in the bottom right of Fig. 8. This gradient is less evident at smaller smoothing levels.

At both test bed sites, the small RMVs found could also be identified while compaction is in progress and roller parameters altered to better compact that region of the cell, (i.e., IC).

An image of the estimated process level that contained all red and orange, that is, is everywhere below the threshold value, would return a solid red credibility plot at the highest smoothing levels.

4.3 Improved Intelligent Compaction and Quality Assurance

The combination of sequential modeling, estimation, and analysis of RMVs as outlined above has the potential to improve the IC and QA of current road construction practice. The computationally efficient methods developed provide a robust estimation and decision making program for modern earthwork compaction roller operators.

The sequential, spatial mixed-effects model addresses the inherent difficulties of RMV data. The utilization of multiple spatial processes in a sequential modeling approach addresses concerns of multiple scales of variation. The coordinate transformation done by centering and scaling the coordinate system addresses the range anisotropy found in the data. The spatial mixed-effects model addresses nonstationarity issues. The state-space formulation of the model corrects for deviations in driving by mapping data locations to a regular grid. The density of the grid of the state-space can be chosen to minimize computation time for sampling while maintaining accurate estimation.

The sequential, spatial backfitting routine provides estimates of the model parameters. The algorithm has been proven to converge and converges in a small amount of time in practice. Utilization of the compact support of spherical covariance matrices to induce sparse matrices speeds this computation time and convergence of the estimation.

The modified multiresolution scale space analysis provides credible regions of the compaction site to identify hard and soft spots. This pseudo-Bayesian approach to image processing provides credible regions in a graphical display that quickly identifies regions of the compaction site that potentially require more compaction. The entire implementation of this modeling, estimation, and analysis is compuationally efficient and can be run on an on-board laptop.

5 Conclusions

The detailed methodology of sequential, spatial modeling and backfitting of RMVs coupled with multiresolution scale space analysis of the resultant estimate images can be utilized to improve IC. By implementing a sequential model structure, the spatial uncertainty in RMV data is used to generate estimates of the true compaction level. The use of spherical covariance matrices speeds computation time of the estimation and the resultant images can be produced at a resolution that provides speed in computation of the multiresolution scale space analysis step. Implementation of such a scheme is time effective and has the potential to improve the rudimentary IC currently utilized.

The utilization of a sequential modeling and estimation approach also provides information on individual layers of compaction unavailable from the raw data. As the roller measures to a depth of several layers of material, the measurement is truly a composite measurement of several layers of the compaction site. The sequential modeling and estimation proposed decomposes the RMV data collected into individual layers of compaction, potentially providing more information on the compaction state of the current layer of compaction than is available from the raw data.

The sequential, spatial mixed-effects model uses spatial, random processes in its additive decomposition. Splines can also replace these random processes. The estimation of a spline is mathematically equivalent to the universal kriging done in this paper, as detailed in Heersink and Furrer (2013). The literature

on splines is extensive and computational feasibility can be maintained (Eilers et al., 1996; Eilers and Marx, 2010; Marx and Eilers, 1998; Wahba, 1990).

Considering the Bayesian nature of the multiresolution scale space analysis, a fully Bayesian approach to the sequential modeling could also be utilized. In such a framework, the RMVs would be modeled as conditionally multivariate Gaussian and suitable priors applied to the modeling terms.

Acknowledgements The roller data was collected and provided by Dr. Mike Mooney as part of the NCHRP 21-09 project of the Transportation Research Board of The National Academies. This work is supported in part by the Swiss National Science Foundation through Project 200021-129782.

References

Breiman L, Friedman JH (1985) Estimating optimal transformations for multiple regression and correlation. Journal of the American Statistical Association 80(391):580–598

Buja A, Hastie T, Tibshirani R (1989) Linear smoothers and additive models. The Annals of Statistics 17:453–510

Eilers PHC, Marx BD (2010) Splines, knots, and penalties. Wiley Interdisciplinary Reviews: Computational Statistics 2(6):637–653, DOI 10.1002/wics.125

Eilers PHC, Rijnmond DM, Marx BD (1996) Flexible smoothing with B-splines and penalties. Statistical Science 11:89–121

Facas N, Mooney MA, Furrer R (2010) Anisotropy in the spatial distribution of roller-measured soil stiffness. Int J Geomech 10(4):129–135, DOI 10.1061/(ASCE)GM.1943-5622.0000053

Facas NW (2009) Variogram properties and anisotropy in the spatial distribution of roller-measured soil stiffness. Master's thesis, Colorado School of Mines

Facas NW, Mooney MA (2011) Characterizing the precision uncertainty in vibratory roller measurement values. Journal of Testing and Evaluation 40:1–9

Furrer R, Sain SR (2009) Spatial model fitting for large datasets with applications to climate and microarray problems. Statistics and Computing 19(2):113–128, DOI http://dx.doi.org/10.1007/s11222-008-9075-x

Heersink DK, Furrer R (2013) Sequential spatial analysis of large datasets with applications to modern earthwork compaction roller measurement values, submitted to *Spatial Statistics*

Holmström L, Pasanen L, Furrer R, Sain SR (2011) Scale space multiresolution analysis of random signals. Computational Statistics & Data Analysis 55(10):2840–2855

Marx BD, Eilers PHC (1998) Direct generalized additive modeling with penalized likelihood. Computational Statistics & Data Analysis 28(2):193–209, URL http://EconPapers.repec.org/RePEc:eee:csdana:v:28:y:1998:i:2:p:193-209

Mooney MA, Rinehart RV, White DJ, Vennapusa PK, Facas NW, Musimbi OM (2010) Intelligent soil compaction systems: NCHRP project 21-09 final report

Petersen D, Erickson M, Roberson R, Siekmeier J (2007) Intelligent soil compaction: geostatistical data analysis and construction specifications. In: Transportation Research Board 86th Annual Meeting

Scherocman J, Rakowski S, Uchiyama K (2007) Intelligent compaction, does it exist? In: Proceedings of the 52 Annual Conference of the Canadian Technical Asphalt Association, Polyscience Publications Inc., URL http://www.amazon.com/Proceedings-Conference-Canadian-Technical-Association/dp/0921317670

Thurner H, Sandström A (2000) Continuous compaction control. In: Proceedings of the European Workshop Compaction of Soils and Granular Materials, pp 237–246

Wahba G (1990) Spline Models for Observational Data. CBMS-NSF Regional Conference Series in Applied Mathematics, Society for Industrial and Applied Mathematics, URL http://books.google.ch/books?id=ScRQJEETs0EC

White DJ, Thompson MJ (2008) Relationships between in situ and roller-integrated compaction measurements for granular soils. Journal of Geotechnical and Geoenvironmental Engineering 134(12):1763–1770, DOI 10.1061/(ASCE)1090-0241(2008)134:12(1763)

Zimmerman DL (1993) Another look at anisotropy in geostatistics. Math Geol 25(4):453–470

Paper V

Spatial Backfitting of Roller Measurement Values from a Florida Test Bed

Daniel K. Heersink, Reinhard Furrer, & Mike A. Mooney

Report, arXiv:1302.4659.

Spatial Backfitting of Roller Measurement Values from a Florida Test Bed

Daniel K. Heersink*[1], Reinhard Furrer[1], and Mike A. Mooney[2]

[1]Institute of Mathematics, University of Zurich, CH-8057 Zurich
[2]Colorado School of Mines, Golden, CO 80401, USA

Modern earthwork compaction rollers collect location and compaction information as they traverse a compaction site. These data are indirectly observed through non-linear measurement operators, inherently multivariate with complex correlation structures, and collected in huge quantities. The nature of such data was investigated at a large, atypically compacted test bed in Florida, USA. Exploratory analysis of this data through detrending and empirical semivariogram estimation is performed. A second analysis using a sequential, spatial backfitting algorithm is used to investigate the importance of driving direction of the roller.

Keywords: Spatial backfitting; sequential modeling; semivariogram estimation; anisotropy

1 Modern Earthwork Compaction

Modern compaction rollers monitor soil properties by observing stiffness characteristics of the soil. A vibrating drum traverses the compaction site at approximately 1m/s, compacting approximately 20cm of material at a time. Common construction practice is to compact several layers of material during the construction of a new road. Each layer is compacted in several passes of the roller until sufficient compaction is achieved.

*To whom correspondence should be addressed. Email: daniel.heersink@gmail.com

Typical construction practice is to compact in segments of road 10–15m wide and 50–100m long. The roller traverses the compaction site in a snaking motion of several adjacent lanes. In practice, there is very little overlap between lanes (Mooney *et al.*, 2010). See Figure 1 for a typical compaction roller manufactured by Ammann.

Figure 1: Ammann roller at work.

1.1 Roller Measurement Values (RMVs)

A typical smooth drum has a diameter of approximately 1m and is approximately 2m long. An on-board sensor and GPS system record measurements that are together termed the roller measurement value (RMV). An individual RMV is an aggregate measure of a bulb of soil extending to a depth of approximately 1m with a diameter of 0.5–0.6m (Facas, 2009).

The physical nature of driving the roller down a lane with its vibrating drum causes other vibrational "wobbling" that remains fairly uniform over the course of the entire lane. Any bias this action produces will therefore be uniform over the entire lane. When the roller turns around and makes another pass down a different lane, the "wobbling" effect may be different though. This will lead to a change in the bias in the transverse direction, but the driving direction should remain unchanged as the new bias will be uniform over that entire lane. This is a cause of potential measurement error found only in the transverse direction.

1.2 Florida Test Bed Data

For a detailed investigation of roller properties, statistical characteristics, etc., a test bed with atypical dimensions was atypically densely compacted. A compaction roller traversed the compaction site in both the x- and y-directions. This Florida dataset consists of 19,145 observations of x- and y-coordinates, soil stiffness (k_s), and lane number in the x-direction driving and 19,975 observations in the y-direction driving. This analysis focuses on the driving direction.

The roller first traversed the compaction site in the x-direction in a snaking fashion, first left-to-right and then back again right-to-left. The roller then traversed the compaction site a second time in the y-direction. The physical limitations of the site prohibited a snaking traversal in the y-direction, so the roller moved from bottom to top only. There are 29 lanes in the x-direction and 27 lanes in the y-direction. Figure 2 is a plot of the RMVs in the x-driving direction and the y-driving direction. Blue values represent high stiffness and red values represent low. An optimally compacted site would be uniformly blue.

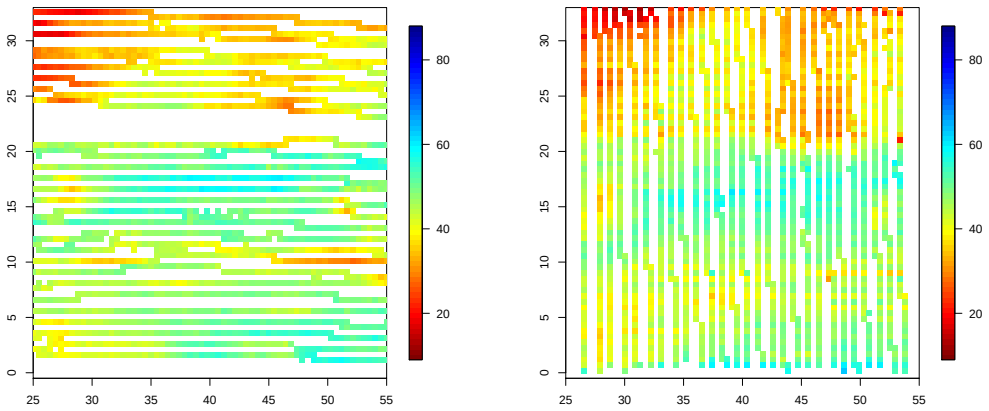

Figure 2: Data from the test bed in Florida, USA. RMVs collected from driving in the x-direction (left) and from driving in the y-direction (right) are depicted.

2 Exploratory Data Analysis

For this analysis, the x- and y-direction driving data are treated as two separate datasets. First, empirical semivariograms of the raw data were calculated using a subsample for computational reasons. These semivariograms exhibit aspects of non-stationarity. See Figure 3 for representative empirical semivariograms of both driving directions.

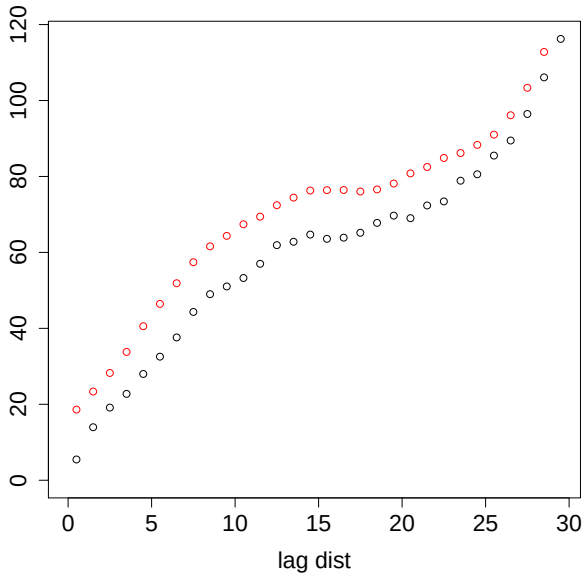

Figure 3: Empirical semivariograms of raw RMV data for x-direction driving (black) and y-direction driving (red).

2.1 Detrending the Data

The raw data exhibits a mean trend that must be removed as a constant mean is required to attain second-order stationarity. By detrending the data, we can remove the mean trend and proceed with the analysis utilizing a second-order stationary spatial process as a model.

2.1.1 Small and Large Scale Variation

Often times spatial data is modeled as

$$y(\boldsymbol{s}) = \mu(\boldsymbol{s}) + \alpha(\boldsymbol{s}) + \varepsilon(\boldsymbol{s}), \tag{1}$$

where $bmu(\boldsymbol{s})$ is the mean structure of the process, $\alpha(\boldsymbol{s})$ is the stochastic dependence structure of the process, and $\varepsilon(\boldsymbol{s})$ is the measurement error. The mean structure is termed large scale variability and the dependence structure is termed small scale variability. What is termed mean structure and what is termed covariance structure is largely discretionary (Cressie, 1993).

2.1.2 Detrending Methods

Assuming model (1), we desire a second-order stationary process $\alpha(\boldsymbol{s})$. Therefore, the data detrending process should leave some structure in the data or all that will be left is the noise process $\varepsilon(\boldsymbol{s})$, which is assumed uncorrelated. We have an assumption of spatial correlation.

The natural first choice for detrending is fitting a linear model: $\mu(\boldsymbol{s}) = \mathbf{X}\boldsymbol{\beta}$. The residuals of the linear model fit can then be used to estimate the semivariogram of the stochastic structure terms $\alpha(\boldsymbol{s}) + \varepsilon(\boldsymbol{s})$. The detrending process used included all cross products of the x- and y- coordinates. That is, for a 4th order polynomial, all products of x and y with a combined power of 4 or less were used. Empirical semivariograms were then calculated on the residuals of the linear model. Using a polynomial detrending of a 5th power generates empirical semivariograms with qualitatively identifiable nugget, partial sill and range parameters. This degree of detrending is desirable as all spatial variation is not lost and a constant mean of the residuals has been attained.

A practical, physical explanation of the linear model parameters is not of importance. The goal of detrending is establishing a constant mean of the residuals, and interpretability of the model parameters is insignificant.

An alternative to linear regression for data detrending is to detrend the data using a nonparametric function. For this analysis, the implementation of local polynomial smoothing known as locally weighted scatterplot smoothing (loess) was used (Cleveland, 1979). The loess smoothing approach is based on a moving window. A polynomial is fit to the data in a window using robust methods. The fitted value is then the predicted response at the middle of the window. The window is then slid over the range of the data, repeating the fitting process as the window moves (Faraway, 2006).

For this analysis, a span of 0.5 was used to reproduce empirical semivariograms like those of the polynomial detrend. This span corresponds to an estimated number of parameters of 13.5. This is approximately equivalent to a polynomial fit of 4th order, making this method comparable to that of a polynomial detrending.

2.2 Fitting to a Model

The empirical semivariograms calculated from the loess detrended data were then fitted to a spherical model with Cressie weights using the `variofit` function in `R`. The spherical model was chosen as the empirical semivariograms seemed to exhibit a linear behavior near the origin. The spherical model also induces sparse matrix structures, helpful for computation. The spherical model is defined as

$$C(h;\boldsymbol{\theta}) = \begin{cases} \theta_0(1 - 1.5(h/\theta_1) + .5(h/\theta_1)^3) & \text{for} \quad h \in [0, \theta_1) \\ 0 & \text{for} \quad h \geq \theta_1 \end{cases}, \tag{2}$$

where θ_0 is the (partial) sill and θ_1 is the range of the spatial process.

Cressie weights were chosen because they are the most commonly used weights for fitting empirical semivariograms to a covariance model. Weighted least squares and generalized least squares require knowing the covariance structure of the semivariogram. While this is possible, it is hard to implement. Cressie (1985) proposed a weighting structure that is a compromise of weighted least squares that is no more difficult to compute than

ordinary least squares.

2.3 Semivariogram Uncertainty

Estimates of total sill, range, and nugget have very large confidence intervals. As the lag distance increases, the confidence interval for the total sill also increases (Nordman and Caragea, 2008). A simple simulation of several random fields with semivariogram parameters chosen to match those of the empirical semivariograms from this study was performed. From these random fields, empirical semivariograms were then calculated. A mean and standard deviation of these semivariograms was then calculated and these were used to calculate pointwise confidence intervals. The estimated confidence bound of the semivariogram starts very small for a lag distance of zero and begins expanding for larger lag distances. This expansion continues for larger lag distances. Decreasing the confidence to 75% does very little to improve the width of the estimated confidence bounds for large lag distances, see Figure 4.

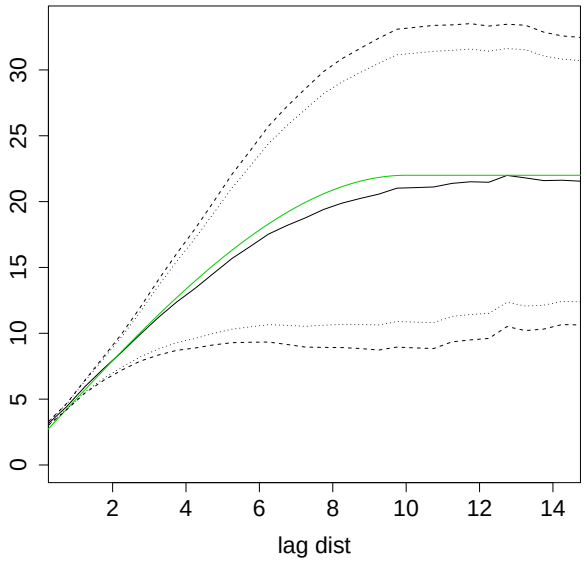

Figure 4: Mean of simulated empirical semivariograms (solid line) and 95% (dashed line) and 75% confidence bands (dotted line). The true spherical semivariogram is depicted in green.

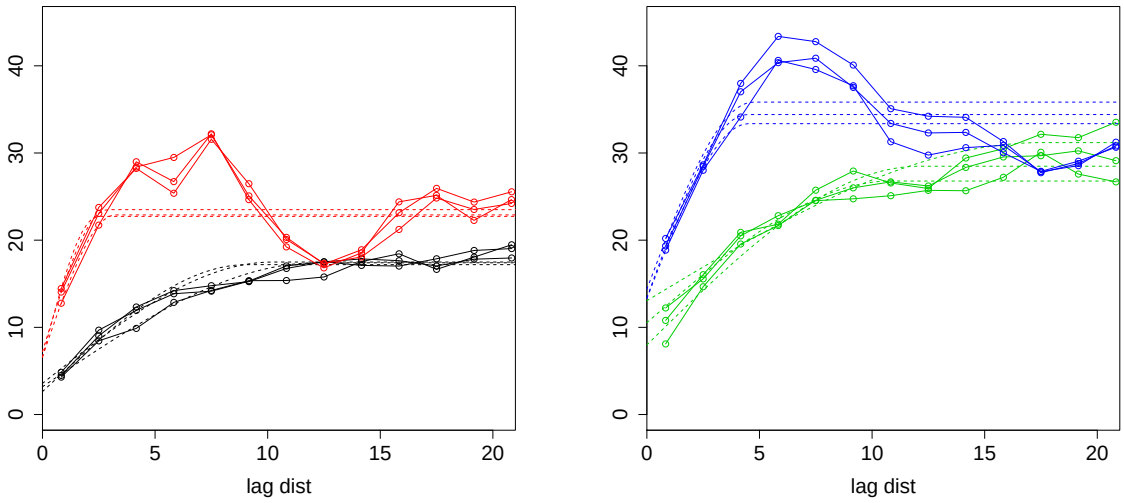

Figure 5: Directional empirical semivariograms and fitted spherical models from three polynomial detrended subsamples of RMVs of the x-driving direction subset (left) and of the y-driving direction subset (right). Dashed lines indicated fitted models, x-directional semivariograms are in black and green and y-directional semivariograms are in red and blue.

2.4 Sampling Concerns

To maintain computational efficiency, the data was subsampled for empirical semivariogram estimation. 10,000 data points were sampled from each of x- and y-direction driving datasets. A loess detrending of each sample was performed. This produced two detrended datasets from which subsamples of 2500, 3500, and 4500 data points were drawn. Directional empirical semivariograms were then calculated in the x- and y-direction to generate a total of twelve empirical semivariograms. These empirical semivariograms were then fit to a spherical model.

There was no discernible difference between the empirical semivariograms within each dataset. Figure 5 depicts the empirical and fitted directional semivariograms of the x-driving direction dataset (left) and y-driving direction (right). Since the sampled directional empirical semivariograms are essentially identical within each dataset, we concluded the subsampling was adequate, i.e. the subsampling produced a representative sample

2.5 Results

The next step is a qualitative analysis of the characteristic semivariogram features. Semivariograms for both driving directions exhibit similar features. For both driving directions, the y-directional semivariograms have a range of 0–5 and the x-directional semivariograms exhibit a range of 9–15.

For the x-driving direction, the total sill for y-directional semivariograms is 22–25, and 15–18 for x-directional. In the y-driving direction, the total sill is 33–36 for y-directional and 26–30 for x-directional semivariograms. Similarly, for x-driving direction, the nugget for y-directional semivariograms is 5–8 and 2–5 for x-directional. For the y-driving direction, the nugget for y-directional semivariograms is 10–15 and 10–12 for x-directional semivariograms, see Table 1 and Figure 5.

Table 1: Directional semivariogram parameters

	x-driving		y-driving	
	x-directional	y-directional	x-directional	y-directional
range	9–15	0–5	9–15	0–5
total sill	15–18	22–25	26–30	33–36
nugget	2–5	5–8	10–12	10–15

3 Anisotropy Concerns

Based on these observations, it is fairly safe to assume that there is no sill or nugget anisotropy. There does appear to be a range anisotropy between the x-directional semivariograms and the y-directional semivariograms. The ratio of the range in the x-direction vs. the y-direction is approximately 5:1. The empirical semivariograms indicate a geometric range anisotropy that can be dealt with by a simple transformation of the data locations, (Zimmerman, 1993).

This geometric range anisotropy can possibly be explained by the compaction process. As the roller traverses the compaction site, it collects data every 10cm in the driving direction. Data is collection in the direction perpendicular to the driving direc-

tion approximately every 1–2m. The vibrating drum is approximately 2m wide, thus the y-directional location of observations in adjacent lanes is 1–2m apart, dependent on the placement of the GPS unit. Also, material is brought into the compaction site via a dump truck and laid down in sections. It is unclear if discontinuities exist on the boundaries of these sections. If they do exist, they could contribute to range anisotropy.

Due to the nature of the driving process, data points are much more closely spaced in the driving direction than they are in the transverse direction. This leads to difficulties estimating the nugget in the transverse direction as the smallest lag distance is on the order of 1–2m. The nugget anisotropy could therefore be explained by a vertical shift of the entire semivariogram caused by a measurement error in the transverse direction. This would essentially be a nugget anisotropy model.

Let the true compaction process be denoted by $Z(x, y)$ and the data we collect be denoted by $Y(x, y) = Z(x, y) + \varepsilon(y)$, where $\varepsilon(y)$ is a measurement error seen only in the y-direction. Then, the semivariogram of the Y process is $\gamma_Y(h_x, h_y) = \text{Var}(Y(x, y) - Y(x + h_x, y + h_y)) = \text{Var}(Z(x, y) - Z(x + h_x, y + h_y) + \varepsilon(y) - \varepsilon(y + h_y)) = \gamma_Z(h_x, h_y) + \gamma_\varepsilon(h_y)$. The x-directional semivariogram is then $\gamma_x(h_x) = \gamma_Y(h_x, 0) = \gamma_Z(h_x, 0)$ and the y-directional semivariogram is $\gamma_y(h_y) = \gamma_Z(0, h_y) + \gamma_\varepsilon(h_y)$. Thus the transverse directional semivariogram is shifted up by the measurement error ε.

4 Driving Direction Investigation

We utilize a state-space formulation to handle unique observation locations. Assume the RMVs can be decomposed into an underlying mean trend dependent on spatial location, driving direction, speed, and vibration amplitude, and a Gaussian spatial random process, (i.e. $\boldsymbol{w} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\alpha}$), where the domain of $\boldsymbol{w}$ is a lattice. Here, $\mathbf{X}$ is a full rank matrix of the fixed effects covariates and $\boldsymbol{\alpha}$ represents an unknown, spatially varying random process. The observed locations of the RMVs are then mapped to the lattice.

Implementing a sequential, spatial mixed-effects model Heersink and Furrer (2013),

we can model the Florida dataset as:

$$\boldsymbol{z}_x = \mathbf{H}_x\mathbf{X}_x\boldsymbol{\beta}_x + \mathbf{H}_x\boldsymbol{\alpha}_x + \boldsymbol{\varepsilon}_x$$
$$\boldsymbol{z}_y = \mathbf{H}_y\mathbf{X}_y\boldsymbol{\beta}_y + c\mathbf{H}_y\boldsymbol{\alpha}_x + \mathbf{H}_y\boldsymbol{\alpha}_y + \boldsymbol{\varepsilon}_y,$$

where $\boldsymbol{\alpha}_x$ and $\boldsymbol{\alpha}_y$ correspond to random variation of the layer of material being compacted during driving in the x- and y-direction, $\mathbf{H}_x$ and $\mathbf{H}_y$ are the operators mapping lattice points to observed locations, and $\boldsymbol{\varepsilon}_x$ and $\boldsymbol{\varepsilon}_y$ represent the measurement error of the sensor. For this analysis, the lattice chosen is an 80×80 grid of points equally spaced on $[25, 55] \times [-.5, 33]$. The size of the grid was chosen to encompass all observation locations.

We also utilize a range anisotropy parameter ρ, given the empirical semivariograms calculated in Section 2.4. The range anisotropy is handled with a transformation of the coordinates. Thus ρ is the ratio of the range in the x-direction to that in the y-direction and the transformation matrix $\mathbf{A}$ is defined as $\mathbf{A} = \text{diag}(1, \rho)$.

As detailed in Heersink and Furrer (2013), any additive term that can be estimated in a mathematically equivalent way as universal kriging can also be included in such a model. Splines are such an additive component that has this mathematical equivalency. The literature on splines is extensive and computational feasibility can be maintained, i.e. Wahba (1990), Eilers *et al.* (1996), Marx and Eilers (1998), Eilers and Marx (2004).

Since there was not a new layer of material added to the compaction site after compacting in the x-direction, the measurements in the y-direction are measurements of the same process as those in the x-direction. Thus, we should expect $c \to 1$ and either $\boldsymbol{\alpha}_y \to \mathbf{0}$ or $\boldsymbol{\alpha}_y \to \boldsymbol{\gamma}_y$, where $\boldsymbol{\gamma}_y$ represents a spatially varying process only in the y-direction, e.g. a process representing the nugget anisotropy discussed in Section 3. Thus we would expect to see an empirical semivariogram of $\boldsymbol{\alpha}_y$ to either have a sill of zero or a very small range. Due to measurement errors, a pure nugget model is not expected.

The Sequential Backfitting Algorithm from Heersink and Furrer (2013) was applied to the data, setting $c = 1$, with $p = 2$, $\mathbf{X}$ is the fixed effects matrix containing all 5th

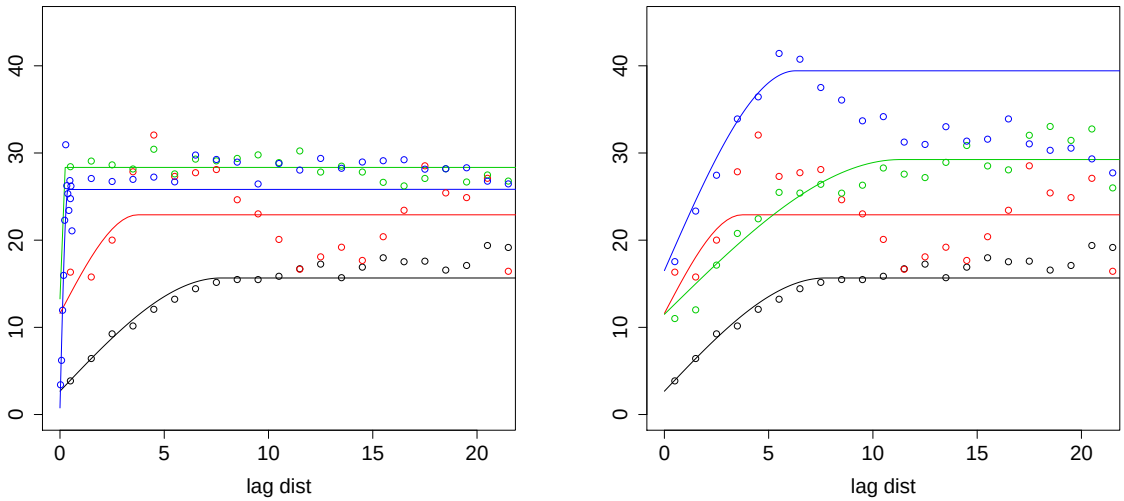

Figure 6: Fitted x-directional semivariograms for $\boldsymbol{\alpha}_x + \boldsymbol{\varepsilon}_x$ (black) and $\boldsymbol{\alpha}_y + \boldsymbol{\varepsilon}_y$ (green) and y-directional semivariograms for for $\boldsymbol{\alpha}_x + \boldsymbol{\varepsilon}_x$ (red) and $\boldsymbol{\alpha}_y + \boldsymbol{\varepsilon}_y$ (blue) for $c = 1$ (left) and $c = 0$ (right). The $c = 0$ plot reproduces the curves from Figure 5, as would be expected.

degree and lower polynomial combinations of the centered and scaled x- and y-direction coordinates of the roller. To create sparse matrix structures and aid in computation a spherical covariance function was assumed, see equation (2).

The semivariogram estimation done in this study is directional. The empirical semivariograms were calculated in the driving direction. Thus, for $\boldsymbol{\alpha}_x$ empirical semivariograms were calculated in the x-direction and in the y-direction for $\boldsymbol{\alpha}_y$.

5 Backfitting Results

The estimated covariance parameters for $\boldsymbol{\alpha}_x$ are $\widehat{\boldsymbol{\theta}}_x = (12.99, 7.72)^\mathsf{T}$ and $\widehat{\boldsymbol{\theta}}_y = (25.08, 0.39)^\mathsf{T}$ for $\boldsymbol{\alpha}_y$, the estimated variances of $\boldsymbol{\varepsilon}_x$ and $\boldsymbol{\varepsilon}_y$ are $\widehat{\sigma}^2_x = 2.68$ and $\widehat{\sigma}^2_y = 0.75$, see Figure 6. The backfitting procedure thus reproduces the empirical x-directional semivariogram from the standard detrending approach.

The range of the $\boldsymbol{\alpha}_y$ process is relatively small, thus there is no evidence to reject the assumption that $\boldsymbol{\alpha}_y \to \boldsymbol{\gamma}_y$ from this analysis. This backfitting analysis thus reconfirms the

existence of a nugget effect in the y-direction. This would imply the "static" rolling done by the roller after compaction was completed is generally truly static and the material is not being actively compacted during this phase of construction.

The backfitting procedure was also run for $c = 0$. As can be seen in the right plot of Figure 6, the calculated semivariograms are reproductions of the standard detrending approach of Section 2.1.2 and the empirical semivariograms found in Figure 5.

References

Cleveland, W. S. (1979). Robust Locally Weighted Regression and Smoothing Scatterplots. *Journal of the American Statistical Association*, **74**, 829–836.

Cressie, N. (1985). Fitting variogram models by weighted least squares. *Mathematical Geology*, **17**, 563–586.

Cressie, N. (1993). *Statistics for spatial data*. Wiley series in probability and mathematical statistics: Applied probability and statistics. J. Wiley.

Eilers, P. H. C. and Marx, B. D. (2004). Splines, knots and penalties. Technical report.

Eilers, P. H. C., Rijnmond, D. M., and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, **11**, 89–121.

Facas, N. W. (2009). Variogram properties and anisotropy in the spatial distribution of roller-measured soil stiffness. Master's thesis, Colorado School of Mines.

Faraway, J. J. (2006). *Extending the linear model with R*. Texts in Statistical Science Series. Chapman & Hall/CRC, Boca Raton, FL.

Heersink, D. K. and Furrer, R. (2013). Sequential spatial analysis of large datasets with applications to modern earthwork compaction roller measurement values. Submitted to *Spatial Statistics*.

Marx, B. D. and Eilers, P. H. C. (1998). Direct generalized additive modeling with penalized likelihood. *Computational Statistics & Data Analysis*, **28**, 193–209.

Mooney, M. A., Rinehart, R. V., White, D. J., Vennapusa, P. K., Facas, N. W., and Musimbi, O. M. (2010). Intelligent soil compaction systems: NCHRP project 21-09 final report.

Nordman, D. J. and Caragea, P. C. (2008). Point and interval estimation of variogram models using spatial empirical likelihood. *Journal of the American Statistical Association*, **103**, 350–361.

Wahba, G. (1990). *Spline Models for Observational Data*. CBMS-NSF Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Mathematics.

Zimmerman, D. L. (1993). Another look at anisotropy in geostatistics. *Mathematical Geology*, **25**, 453–470.